DOCUMENT RESUME

ED 177 982 HE 011 879

TITLE The Council of Ontario Universities Experimental Achievement Testing Programme. Summary Report.
INSTITUTION Council of Ontario Universities, Toronto. Research Div.
PUB DATE Sep 79
NOTE 164p.
AVAILABLE FROM Council of Ontario Universities, 130 St. George Street, Suite 8039, Toronto, Ontario M5S 2T4, Canada

EDRS PRICE MF01/PC07 Plus Postage.
DESCRIPTORS *Ability Identification; *Achievement Tests; *College Entrance Examinations; *College Freshmen; *Evaluation Methods; Foreign Countries; Grades (Scholastic); Higher Education; Predictive Ability (Testing); *Predictive Measurement; Predictive Validity; Predictor Variables; Testing; Test Results
IDENTIFIERS Canada

ABSTRACT

To evaluate the usefulness of standardized achievement tests in assessing pre-university academic achievement and aptitude, the Council of Ontario Universities administered tests to incoming freshmen at four Ontario universities, and also examined their high school and subsequent university performance. The tests were designed to measure achievement in English and Mathematics; a writing sample was also taken. The tests used were evaluated for technical quality and validity. It was generally concluded that the tests utilized were found to have generally satisfactory technical qualities, and their utility in evaluating present ability and predicting future performance was comparable to those found in similar studies, that these tests would be useful for selection decisions in the absence of data on students' previous year performance, and that the addition of test scores to students' high school performance data would add only slightly to the efficiency of student selection data. It was further concluded that high school marks, despite their deficiencies, remain the best predictors of university success. Appendices include tables that summarize the findings. (Author/PHR)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED177982

ISBN 088799-115-7

THE COUNCIL OF ONTARIO UNIVERSITIES EXPERIMENTAL ACHIEVEMENT TESTING PROGRAMME

SUMMARY REPORT

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Council of Ontario Universities

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

HE 011 229

Council of Ontario Universities
Conseil des Universités de l'Ontario
130 St. George Street
Suite 8039
Toronto, Ontario, M5S 2T4

September, 1979

79-3

This report was prepared by the Research Division of the Council of Ontario Universities. Requests for further information should be directed to J. T. O'Mahony, Research Associate.

TABLE OF CONTENTS

		Page
	LIST OF TABLES	iii
	ABSTRACT	vi
1.0	INTRODUCTION	1
2.0	RESEARCH QUESTIONS	4
	2.1 Achievement Testing in Perspective	4
	2.2 Scope and Limitations of Study	9
3.0	PRE-UNIVERSITY ACADEMIC PERFORMANCE MEASURES	20
	3.1 Sources of Data	20
	3.2 Achievement Tests	22
	3.2.1 The Role of Testing	22
	3.2.2 The Quality of Academic Performance Measures	24
	3.2.3 Achievement Testing in Ontario	30
	3.3 CTEL and CMPT	31
	3.3.1 COU Fall 1975 Testing Programme	31
	3.3.2 Technical Quality of CTEL and CMPT	39
	3.3.3 Validity Evidence for CTEL and CMPT	43
	3.4 RCLAT and the Writing Test	52
	3.4.1 COU Fall 1976 Testing Programme	52
	3.4.2 Technical Quality of RCLAT and the Writing Test	57
	3.4.3 Validity data for RCLAT and Writing Test	61
4.0	ANALYSES	65
	4.1 Introduction to Analyses	65
	4.2 Alternate Academic Performance Measures	71
	4.2.1 Interim Final Year 5 Marks	71
	4.2.2 Comparison of First Year University Credit Weighted Average from University Data Element with Average Computed from Course Marks	74

		Page
4.3	Validation of Achievement Test and Year 5 Measures for Student Classification Decisions	76
4.3.1	Selection	78
4.3.2	Placement	104
4.4	Variations in High School Marking Standards	111
4.4.1	The Problem	111
4.4.2	Methodology	111
4.4.3	Results	120
5.0	SUMMARY OF FINDINGS	123
	Appendix A1	131
	Appendix A2	132
	Appendix A3	133
	Appendix A4	134
	Appendix A5	135

LIST OF TABLES

Table		Page
1	COU 1975: Programme of Registration for CTEL, CMPT Achievement Test Writers and all First Year Registrants at Brock, McMaster, Queen's and Waterloo, and for all Ontario Universities	34
2	COU 1975: Educational Source of all First Year Registrants and Achievement Test Volunteers at Brock, McMaster, Queen's and Waterloo, and of all Ontario universities	35
3	CTEL: Internal Characteristics - Total Sample	40
4	CMPT: Internal Characteristics	41
5	Correlations of CTEL Scores with High School and University Subject Averages	49
6	Correlations of CMPT Scores with High School and University Course Subject Averages	51
7	COU 1976: Programme of Registration for RCLAT and Essay Test Writers; All First Year Registrants at Carleton and McMaster, and all First Year Registrants at Ontario Universities	54
8	COU 1976: Educational Source of RCLAT and Essay Test Writers, All First Year Registrants at Carleton and McMaster, and All First Year Registrants at Ontario Universities	56
9	Internal Characteristics of the Reading Comprehension and Language Achievement Test (English) and the Writing Test	58
10	Correlation Coefficients of RCLAT and Writing Test Scores with other Academic Performance Measures	62
11	Relationships between Interim and Final Year 5 Averages for Direct Entrants in Fall '75 to McMaster, Queen's and Waterloo	72
12	Relationships between University Sessional Credit Weighted Averages and those Computed from Course Marks for Direct Year 5 Entrants in 1975	75

Table		Page
13.1	Means, Standard Deviations & Correlations for Year 5 Average (1975-76) and University First Year Programme Average	82
13.2	Means, Standard Deviations & Correlations for CTEL, Year 5 Average (1975-76) and University First Year Average	83
13.3	Simple and Multiple Correlation Coefficients for CMPT and Year 5 Average with First Year University Final Programme Average (1975-76)	84 & 85
14.1	Means, Standard Deviations and Correlations for RCLAT, Year 5 Average (1976-77) and University First Year Programme Average	86
14.2	Means, Standard Deviations and Correlations for RCLAT, Essay, Year 5 Average (1976-77) and University First Year Programme Average	87
15.1	Means, Standard Deviations & Correlations for CTEL, Year 5 English (1975-76) and University First Year Subject Averages	100
15.2	Means, Standard Deviations and Correlations for RCLAT, Year 5 English Average (1976-77) and University First Year Subject Averages	101
15.3	Means, Standard Deviations and Correlations for RCLAT, Essay Scores, Year 5 English Average (1976-77) and University First Year Subject Averages	103
16	Means, Standard Deviations and Correlations for CMPT Subtest Scores, Year 5 Mathematics Averages and University First Year Mathematics Average (1975-76)	105
17.1	Summary of Regression Analyses of First Year Average (1975-76) Prediction for McMaster Registrants from Ten High Schools who wrote CMPT-1	116
17.2	Summary of Regression Analyses of First Year Mathematics Average (1975-76) Prediction for the McMaster Registrants from Ten High Schools who wrote CMPT-1 (Algebra)	117

Table		Page
18.1	Summary of Regression Analyses of First Year Average (1976-77) Prediction for McMaster Registrants from Ten High Schools who wrote RCLAT	118
18.2	Summary of Regression Analyses of First Year English Average (1976-77) Prediction for McMaster Registrants from Ten High Schools who wrote RCLAT	119

ABSTRACT

In June 1975, the Council of Ontario Universities initiated a programme of experimental achievement testing. It was felt by many within the Ontario university community that too many students entering the universities were insufficiently prepared in English language and mathematics. Furthermore, the diversity in the academic backgrounds of entering students and in the standards used for evaluating their levels of academic achievement had reached the point where the universities were finding difficulty in making informed admissions and placement decisions, and in designing appropriate first year courses. A proposal was developed for the experimental administration of standardized achievement tests in English and mathematics to groups of entering freshmen at four Ontario universities in the fall of 1975.

Tests were obtained from the Association of Universities and Colleges of Canada and administered in September 1975 with the assistance of funds made available by the Ontario Ministry of Colleges and Universities (MCU). Analysis of results was directed by a steering committee drawing its membership from COU, MCU and the Ministry of Education. The Committee agreed that an extension of the experimental testing programme was worthwhile and an objective test of English language achievement was administered to most first year students in three Ontario universities in September 1976. In addition, a large sample of students at one of the universities also completed a writing test.

In addition to test data, data on high school and subsequent first year university academic performance for all students registered in first year programmes at the participating universities in 1975 and 1976 were also acquired. The present report summarizes the direct findings from both experimental test administrations and the additional analyses performed by COU on the broader data base provided by high school and university performance measures.

The major objective of these testing projects was to evaluate the usefulness of standardized achievement tests for assessing pre-university academic achievement and aptitude, for admission selection or for post-admission diagnostic and placement purposes.

These experimental test administrations and the analysis of the results were hampered by problems of experimental design

(particularly sampling), the lack of suitable criteria for estimating the validity of test results, problems with the quality and consistency of reference data, and limitations of resources. Because of these difficulties, firm conclusions could not be drawn in many areas of the analysis. Nevertheless, certain general conclusions can be made.

The tests utilized were found to have generally satisfactory technical qualities, and their utility in evaluating present ability and predicting future performance was comparable to those found in similar studies. These tests would be useful for selection decisions in the absence of data on students' Year 5 (Grade 13) performance, since the test scores correlate well with Year 5 marks and first-year university marks. Their ability to predict university performance is not as high, however, as that of Year 5 marks. The addition of test scores to high school performance data would provide only slight gains in the efficiency of student selection. Available data showed substantial variability in marking standards of various high schools, but the additional information provided by a single test score was not found to correct for this variation to any great extent. Thus high school marks, despite their deficiencies, remain the best predictors of university success. Of the tests administered in these projects, the highest levels of predictive validity were achieved by the mathematics test (CMPT). Among the English tests, the Writing Test had marginally higher predictive validity than the CTEL and the RCLAT. The Writing Test's reliability, however, was judged to be less than for the objective tests given the manner of its administration.

The design of the achievement testing projects did not permit conclusions to be drawn on the usefulness of the tests for post-admission decisions on placing students in appropriate types and levels of instruction.

The projects provided much useful information on the technical qualities of the tests used, and observations have been made both on their strengths and areas of desirable improvement. Also, the participating universities and COU research staff gained much valuable experience with large-scale test administration, which could be useful in designing future experimentation with or practical application of achievement tests.

The data obtained from the 1975 and 1976 test administrations do not permit any overall observations to be made on aggregate levels of student achievement, because of lack of comparability of the groups tested with other available reference populations. The participating institutions, however, did obtain benchmark data which (with some limitations) could be useful for comparison in any future administrations of these tests.

1.0 INTRODUCTION

In June 1975, the Council of Ontario Universities initiated a programme of experimental achievement testing. It was felt by many within the Ontario university community that too many students entering the universities were insufficiently prepared in English language and mathematics. Furthermore, the diversity in the academic backgrounds of entering students and in the standards used for evaluating their levels of academic achievement had reached the point where the universities were finding difficulty in making informed admissions and placement decisions, and in designing appropriate first year courses. A proposal was developed by the Executive Committee of COU for the experimental administration of standardized achievement tests in English and mathematics to groups of entering freshmen at four Ontario universities in the fall of 1975. The main objectives were to obtain baseline data on the level of students' performance in English and mathematics and to explore the feasibility of administering such tests in future years as post-admission measures of achievement.

Tests were obtained from the Association of Universities and Colleges of Canada and administered in September 1975 with the assistance of funds made available by the Ontario Ministry of Colleges and Universities (MCU). Analysis of results was directed by a steering committee drawing its membership from COU, MCU and the Ministry of Education. The Committee agreed that an extension of the experimental testing programme was worthwhile and an objective test of English language achievement was

administered to most first year students in three Ontario universities in September 1976. In addition, a large sample of students at one of the universities also completed a writing test.

Specialists in testing were contracted to analyse the test results from both the 1975 and 1976 COU experimental administrations. Their findings are available from Council on request.[1]

In addition to test data, data on high school and subsequent first year university academic performance were also acquired by COU for all students registered in first year programmes at the participating universities in 1975 and 1976. The present report summarizes the direct findings from both experimental test administrations and the additional analyses performed by COU on the broader data base provided by high school and university performance measures.

Section 2.0 is devoted to a discussion of the concerns which led to the COU Experimental Achievement Testing Programme and other similar studies. It further examines the contribution that the COU studies are capable of making, given their limitations, in answering some of the questions raised. In particular, it deals with the roles that standardized achievement tests might play in meeting present policy and operational needs in university admissions.

---

[1] Vincent D'Oyley et al Technical Note on the Canadian Test of English Language (CTEL); Vincent D'Oyley et al The Canadian Mathematics Placement Test (CMPT) Technical Report; Ross E. Traub et al Technical Report on the Experimental Administration of the Reading Comprehension and Language Achievement Test (English) for the Council of Ontario Universities; Ross E. Traub et al Technical Report on the Experimental Administration of the Writing Test for the Council of Ontario Universities

Before the practical utility of tests for decision-making can be examined, some mention must be made of their technical qualities as test instruments. Section 3.0 summarizes the more salient findings and provides a brief commentary on concepts and terminology which are indispensable in interpreting the material presented.

In Section 4.0 the questions raised earlier as to achievement test utility in university admissions and placement decision-making are examined in the light of analyses of data acquired from three sources: the experimental administrations themselves; from student records containing biographical and high school marks data obtained from the Ontario Universities Application Centre; and from university student records supplied by those universities which participated in the experimental testing programme. The principal analyses attempted concern: the relationships between alternate measures of academic performance, for example interim and final year 5 marks; the validation of test and other measures for a variety of student classification decisions; and analyses of variations in high school marking standards.

Finally, the findings of the COU Achievement Testing Programme and subsequent data analyses are summarized in Section 5.0, together with a discussion of their implications for future Ontario university admissions and placement practices.

2.0 RESEARCH QUESTIONS

2.1 Achievement Testing in Perspective

A number of serious economic and demographic circumstances preoccupy the Ontario university community at the present time. The simultaneous impact of declining enrolments, governmental fiscal restraint, high unemployment, and unabated inflation is all too familiar. Another not entirely unrelated concern has continued less prominently for some time. Under a variety of guises, dissatisfaction has been been expressed with the level of preparation in first language and quantitative skills of a significant proportion of students entering Ontario universities. Other complaints have been made about the difficulty in assessing the probable university performance of prospective students in particular programmes on the basis of currently available data, and the suspicion has been voiced that variations among high schools in their marking standards may unfairly prejudice the chances of admission to limited enrolment programmes (and perhaps the gaining of scholarships) for applicants from schools with more severe marking policies.

The major problems referred to first above are not unrelated to the questions pertaining to admissions assessment with which this report is primarily concerned. It might be thought that enrolment decline would tend to de-emphasize admissions questions since institutions could now adopt something approaching an open admissions policy.

This view, however, overlooks several factors. First, even though the demand for places in the Ontario university system as a whole may be expected to decline, the competition for admission to a number of programmes has increased, chiefly one supposes as a result of job market conditions and of students' perceptions that such programmes provide an entrée into an occupation. For these programmes at least, the need for equitable admissions criteria and reliable and efficient performance predictors continues in importance. The situation here, however, is somewhat simplified, in that increased competition for admission to a programme will, in general, raise the pre-university academic performance criterion cut-off to a level where the risk of admitting students incapable of coping academically is reduced. There remain, however, such questions as what is the most equitable ranking measure for that programme (subject prerequisites and levels), what can be done to improve the comparability of high school marks and how should applicants not coming from Ontario year 5 (and thus having no year 5 record) be evaluated? At the other end of the scale are those programmes experiencing difficulty in attracting their accustomed student intake. Inevitably admissions officers will be faced with applicants having attained lower levels of academic achievement and for whom not all programmes/courses may be suitable. More careful consideration has to be given to an applicant's (or registrant's) strengths and weaknesses in specific subject and skill areas in order to make placement decisions in programmes or

course sections (assuming differential preparation), to give guidance in the student's selection from the options open to him/her, and to diagnose serious shortcomings which would likely handicap a student's normal progress through university. The emphasis here, then, has shifted from the equitable ranking of undoubtedly qualified applicants for limited enrolment programmes to the matching of applicant potential with programme standards and prerequisites for programmes for which the global minimum admission criterion is the only constraint on entry. The shift is from a concern with equity to the avoidance of wastage and student frustration. In either case the effectiveness of the selection or placement decision rests on the ability to accurately predict an applicant's (or registrant's) subsequent university performance from the information - principally academic - available.

In passing, some comment should be made as to the role of the minimum admission criterion applying to Ontario universities as a whole. An average final mark of 60% on the best six year 5 credits of at least six credits passed, is generally required at present for students applying to Ontario universities directly from year 5. It might be asked why the figure of 60% rather than some other value or indeed why any stipulated average requirement, at all, is necessary. *De facto* entrance standards are set in any event for limited enrolment programmes by the process of competition for places. The minimum,

then, only excludes applicants to relatively open programmes where future economic benefits are less apparent than might be the case with professional programmes. As such, they probably would not attract many high school students with less than 60% average, unless strongly motivated. Also, since there is no evidence of an absolute cut-off point which will discriminate between those who can and those who cannot benefit from a university education (and indeed high school marks themselves are not absolute or standardized measures), the selection effect of the cut-off is somewhat arbitrary for borderline cases.

While granting that the value of 60% does not in itself correspond to any absolute minimal criterion of worthiness for university study, it does, however, serve as one of the reference points which secondary school markers can use when assigning grades (the other two being 50% for a pass and 100% for perfect score), which may possibly serve to standardize high school marks to some extent. One assumes that any change in the value (60%) would merely lead to adaptation by high schools. It also ensures that something more than a token effort is made by prospective university entrants, in a core of year 5 subjects, thus imposing a significant barrier to non-serious applicants. For university admissions officers, a student's proximity to this lower reference point acts as a warning that greater attention be given to other predictors of university performance and more care be taken in placement decisions.

The foregoing should demonstrate that the exact value for the minimum year 5 average criterion is not an appropriate object of empirical research. Its significance resides in its role as a known point of reference. Any change in its value would undoubtedly lead to an adaptive recalibration of high school mark assignments, and the same students would be included or excluded under the new criterion as were under the old. Standardized achievement tests do not suffer from this adaptive nature and thus allow comparisons both over time and cross-sectionally.

The problems in university admissions assessment alluded to thus far have been essentially old problems of academic prediction compounded by new demographic and economic circumstances. The perceived parallel phenomenon of decline in the so-called "basic skills" is causally unrelated to current economic pressures.[2] It appears to be part of a longer term trend of international dimensions and complex origins. It is not our intention to examine this question except to remark that standardized achievement testing provides the only reliable monitoring instrument on which to base remedial policy decisions such as the greater emphasis on language, and mathematics skills in secondary school curricula. Whatever the causes, a wider variation in first language and mathematical

---

[2]College Entrance Examination Board. Advisory Panel on Scholastic Aptitude Test Score Decline, On Further Examination: Report of the Advisory Panel on Scholastic Aptitude Test Score Decline. College Entrance Examination Board (New York, 1977) pp. 44-48

competencies must now be accommodated by universities and the planning for such modification could well benefit from the availability of aggregate data on levels of achievement to be expected from the various categories of incoming students.

2.2 Scope and Limitations of Study

The preceding section has provided a brief discussion of some of the issues connected with admissions assessment for Ontario universities in the late 1970's. In summary, the questions raised above call for baseline data on:

a) aggregate standards of preparation in basic skills - for monitoring the effects of curriculum change in secondary schools and for planning remedial programmes in universities.

b) improved means of assessing an individual student's current achievement and potential academic performance in university to assist in

- the equitable selection of candidates for limited enrolment programmes, and
- post admission decisions for placement, guidance and assignment to remedial programmes.

The purpose of the COU achievement testing projects is to examine the

extent to which data derived from trial administrations of achievement tests in English and mathematics can fill some of these gaps. A further objective is that of gaining some practical experience with the large scale administration of such tests.

The problems raised thus far are open to a variety of possible solutions, ranging from the reinstatement of province wide annual testing in each subject area, through university-set subject entrance examinations or test batteries, to minor modifications of the present system by the addition of supplementary testing for a subset of applicants. Even the adoption of an open admissions policy would not dispense with the need for admissions assessment for purposes of counselling, and for ranking applicants to limited enrolment programmes. There appears to be little support from any group, however, for a return to large scale comprehensive subject examinations either provincially administered or set by the universities themselves. Such measures would be costly while holding out little promise of a commensurate improvement over the existing system. For the purpose of predicting first year university average grades, recent research[3] has shown the "departmentals" were no better in predicting university performance than today's school-generated year 5 final subject averages. The only advantages accruing from the use of

---

[3] R. Traub, et al, Secondary-Postsecondary Interface Project II: Nature of Students, Vol. I, 1976 p. 313

common examinations for all applicants is their lack of bias with respect to marking standards and the comparability in subject matter tested.

The disadvantages, including the steering effect that common examinations would have on high school curricula, the fact that they take no account of non-examination evaluation and are a unique assessment rather than being the result of a year of convergent evaluation, would all appear to outweigh the advantage of standardization. If the latter could somehow be achieved without recourse to the cumbersome apparatus of province-wide subject examinations, then most of the criteria for an adequate academic performance measure for university-bound students would be met.

It is in this context that interest in limited achievement testing arose - limited in the sense of not examining every subject area specialization. If standardized achievement tests could be developed which both monitored basic skills and at the same time served to reduce biases caused by variations in school marking standards, then this would be a most satisfactory solution. The increment in predictive validity which must ensue from the use of achievement test scores together with high school marks would be a welcome bonus, as would the ability to apply a common yardstick to both direct secondary school applicants and the much smaller but growing non-secondary component.

In the fall of 1975 and 1976, the Council of Ontario Universities administered achievement tests to large samples of first year university students. In addition to test scores, further individualized data on high school and end of session university achievement were subsequently acquired for the members of these cohorts. The result of the foregoing data collection activities was the creation of computer files containing high school test score and university data for each student - both academic performance and some biographical variables being included.

Before outlining some theoretical and empirical limitations of the COU studies, mention should be made of the relationship of these studies to the INTERFACE projects, in particular the University Records Survey. In many ways the COU achievement testing projects are complementary to the University Records Survey, while having a much different design. The University Records Survey, drawing as it did a stratified random sample from the schools supplying first year registrants to eleven Ontario universities in the fall of 1975, should be more representative of Ontario as a whole than either the 1975 or 1976 COU sample, which consisted of first year registrants at four universities in 1975 and three in 1976. These latter samples, however, were virtually the total populations entering those universities, whereas the University Records Survey's sampling scheme resulted in

approximately a 5% sample of the corresponding cohorts. The larger COU sample size allows a finer breakdown of the sample into more homogeneous groups, e.g. by sex, educational source, programme of registration or course, while maintaining a high level of statistical precision. The major difference between the two studies as far as the academic performance variables measured, was that in addition to the high school and university course marks obtained for each registrant in the URS, the COU studies also included achievement test scores, thus enabling comparisons of the relative and differential efficiencies of high school marks and achievement test scores for predicting first year university marks and in compensating for high school marking biases. Predictor information was also obtained in the COU studies for students who registered (and in many cases were tested) and subsequently withdrew or failed, without being assigned any representative academic performance measure for that session.

Apart from these differences the two studies suffer from many of the same limitations. These will be discussed at greater length when specific tests and analyses are described. For the moment the following general caveats should be borne in mind.

- The students tested in 1975 and 1976 for the COU achievement testing projects were <u>not representative</u>

of first year registrants for Ontario as a whole in those sessions, nor indeed in 1975 were they representative of all first year programmes at the participating universities. This fact limits the usefulness of the aggregate test statistics obtained, for purposes of establishing norms of tested achievement which can be of wide application. It should also be borne in mind that only registrants are included in the studies mentioned and thus any generalization of findings to the applicant pool (even for the participating universities) or to the potential applicant pool (year 5 graduates) is unwarranted.

- There is reason to believe that the administration of the achievement tests was not uniform from university to university in the same year. As to the motivation to perform well on the tests, the students had little to gain as a result of the testing process. This fact, allied to the timing of test administrations coming as they did amidst the confusion of the first year registration period, could well lead to under-achievement for the COU samples. Unlike the 1976 procedure, the 1975 tests were not mandatory for all first year registrants and a certain amount of non-random self-selection took place.

Turning now to the prediction of university "success" and to what can be legitimately inferred from the COU studies, the following observations

are worthy of note. To the extent that prediction parameters estimated from those studies are to be used for selection for limited enrolment programmes, the target population of interest is the future applicant pool for those programmes. The observed population, i.e. that for which we have both predictor (test and/or year 5 data) and criterion information (first year university marks), has been selected in a number of ways from the potential applicant pool:

- the self-selection involved in a student applying to that particular programme;
- the admissions selection resulting in an offer of admission to the student;
- the self-selection resulting in acceptance or rejection of the offer and subsequent registration in the programme;
- self-selection of voluntary withdrawal before an evaluation has been made of the student's performance in the programme;
- other forms of institutional selection which result in missing criterion information, e.g. in some programmes only students who have attained standing for the year are assigned a weighted average or programme rank. Thus the performance of incompletions and failures is not recorded in the same way as that of successful registrants.

It is unlikely that the various selection processes outlined above are random with respect to academic ability. Thus the net result of such a sequence of selections is that the sample observed from examining the end of session university records of first year registrants in any programme consists of the more successful members of the population of interest. Extra care has to be exercised when interpreting statistical estimates based on this selected sample, particularly with regard to the interpretation of correlation coefficients between predictor and criterion variables. The values obtained from the selected sample will underestimate the unrestricted population values, the degree of underestimation increasing with the degree of selection for the programme concerned. This theoretical limitation of such prediction studies is largely unavoidable since it is clearly impractical to randomly select applicants from the applicant pool (random with respect to their academic performance to date), and then randomly assign them to programmes, in order to obtain a complete statistical design. This limitation can be overcome given further assumptions about the applicant pool; crude adjustments to the correlation coefficients can be made to approximate the unselected target population value more closely. However, the last mentioned form of selection, resulting as it does in selection on the criterion variable, will lead to biases in all parameters estimated, including the standard error of the estimate. For this reason the sessional academic performance data

element found in a student's academic record, though it bears the stamp of university approval, may not always be the best criterion to use in prediction studies. If institutionally assigned sessional averages are missing from student records in substantial numbers whereas individual course records are available and relatively complete, then a criterion based on averaging course marks is clearly preferable for the statistical estimation of prediction parameters for the target population as a whole.

The preceding description of sample and data limitations is not intended as a comprehensive treatment, but serves rather to illustrate some of the dangers inherent in too simplistic an interpretation of statistical results. These and further caveats and qualifications will be discussed more fully in conjunction with particular analyses.

The limitations alluded to thus far have been technical in nature. Far more profound and contentious issues are raised by the elevation of first year grades or grade averages to the status of the criterion measure of an applicant's worthiness to undertake university studies, either in general or in a specific programme. Whereas predictor measures may acquire respectability empirically through their efficiency as predictors of the criterion measure, the latter must be justified on external grounds, ultimately as corresponding in a sensible way to valued outcomes. In this regard first year academic performance as measured by course grades or programme

average may be carrying too heavy a load. Briefly some of the concerns about this criterion are:

- mainly as a measure of quantified academic success at university, first year results are somewhat premature, in that final graduation averages are, to judge from a number of studies[4], less predictable from pre-university academic performance than are first year averages;

- as a concurrent criterion measure of academic achievement at the time of admission, final first year university marks come too late. A year of further study has intervened;

- similary, longer term economic and social status criteria of success have been found to be only weakly related to the level of formal academic achievement in university, although the credentials acquired as a result of programme completion are important determinants of subsequent benefits;

- only narrow academic achievement is encompassed by the criterion, ignoring the other less tangible benefits

---

[4] Several such studies are discussed in: American College Testing Program, Research and Development Division, Assessing Students on the way to College: Technical Report for the ACT Assessment Program, Volume 1. (Iowa City, Iowa: The American College Testing Program, 1973) pp 134-135 and pp 145-146

and learning experiences of university life, and yet even as a measure of academic achievement, the criterion is contaminated with non-academic factors such as hand-writing, examination technique, verbal ability, diligence and so on;

- university examinations are by no means perfectly "reliable", thus the criterion measure is imprecise to some extent;

- in the absence of coordinated evaluation mechanisms between universities and between programmes within universities, it is unreasonable to expect uniformity in marking standards. The criterion, therefore, is not independent of the evaluation unit, i.e. the course or programme in which the student is assessed.

Despite the foregoing weaknesses identified in the criterion, it remains the most commonly used measure for prediction studies and certainly does provide information relevant to the assessment of university achievement.

3.0 PRE-UNIVERSITY ACADEMIC PERFORMANCE MEASURES

3.1 Sources of Data

In order to answer the questions raised in the preceding chapter, three types of academic performance data were acquired for the target cohorts of students, namely;

- achievement test scores
- high school course marks
- first year university course marks

The most convenient source of pertinent high school data was the Ontario Universities' Application Centre (OUAC) which collects and maintains records of "new" undergraduate applicants to the Ontario university system, mainly for purposes of coordinating the communications between applicants and institutions, and also for monitoring aggregate trends in applications. Since universities make use of these applicant records to assess candidates for admission, they contain much of the relevant pre-university academic performance history of an applicant, together with unique individual and institutional identifiers (which are necessary if high school, test and university records from separate sources are to be matched and concatenated to construct a complete record for each student).

First year university data were acquired directly from each

participating institution. The problems encountered in accommodating the diversity in record formats, codes and classification schemes in order to obtain some comparable measures and description of university performance, were substantial. Only the fact that this source was not uniform in the content and definitions of the data elements of research interest needs be noted.

Current samples of Ontario students' scores on widely recognized and appropriate achievement tests were not available from institutional sources, thus necessitating the initiation of an experimental achievement testing programme by COU in the fall of 1975. Competency in both English and mathematics was tested in this administration. The decision was subsequently made to continue the testing programme in the following year (1976), this time employing the English tests used in the Interface Study. No mathematics tests were included in the COU fall 1976 test administration.

The history, development and detailed technical discussion of the two experimental test administrations are both interesting and instructive. However, to treat the test descriptions and analyses fully in the main body of this report would be to digress from our principal objective of investigating the uses of achievement tests in mitigating a number of problems related to selection and student preparedness.

Furthermore, since considerable technical expertise is necessary for achievement test selection, administration, scoring, and item analyses, these components were subcontracted to test specialists in both 1975 and 1976 and were reported separately. To this extent the two test administrations and ensuing technical reports stand alone and may be used as references for other studies and test development quite apart from their application to the questions at issue here. Therefore, only findings having a direct bearing upon these questions will receive attention in the main text; the more detailed reporting of each test administration is available from COU on request. Having introduced the sources of the data available for analysis, we now proceed to discuss each in turn and their significance for our purposes. In doing so it is almost impossible to avoid some excursions into the realm of the psychometricians. An attempt has been made to explain possibly unfamiliar concepts and terminology in as straightforward and non-statistical a manner as can be managed without trivializing the topic unduly.

3.2 Achievement Tests

3.2.1 The Role of Testing

The kinds of educational tests which concern us here are evaluation instruments designed to provide information relevant to

one or more decisions made by, or on behalf of, individual students or populations of students.

Such tests are not designed to measure skills, aptitudes or educational achievement exhaustively, even where this would be feasible, since the costs of such evaluation would be prohibitive for all but the most trivial attainments. These high costs arise from the context in which the testing takes place, i.e. where large numbers of students must be evaluated in a short period of time with limited resources available for the purpose. The tests therefore are designed to maximize the information obtained while paring the resources required (administration time and costs, scoring costs, etc.) to a minimum.

The object of the testing is to gain information about the level of some real-world performance, the knowledge of which will be used in decisions made by or for the individual or group tested. This performance is termed the PERFORMANCE CRITERION for the test. Some performance criteria are sufficiently simple to allow quantification into a single measure which varies in a way consistent with intuitive notions about the behaviour of the criterion. Such a measure is termed the CRITERION MEASURE for the test. For example, admissions officers in universities might be willing to adopt (or define) first year Grade Point Average (GPA) as an acceptable criterion measure of "success" in university and hence

useful for making admissions decisions. Being unable to observe an applicant's GPA directly (since the applicant has yet to enter university) the admissions officers must resort to indirect measures such as aptitude or achievement test scores, or high school marks. Past experience will have demonstrated that knowledge of these indirect measures reduces the uncertainty (i.e. provides information) concerning the actual GPA value that a student obtaining that test or high school average will ultimately obtain at the end of his first year. In the absence, then, of evidence to the contrary the admissions officers' most sensible strategy would be to assume the stability over time of the relationship between predictors (aptitude test etc.) and the criterion measures.

3.2.2 The Quality of Academic Performance Measures

As a general observation, educational tests and examinations provide information for decisions affecting the sequence of learning activities of students. These decisions, by or on behalf of a student, will usually be based upon his/her achievement, aptitude or ability level in skill or knowledge areas important for subsequent learning. For any but the most trivial of learning areas it will be either impossible or impractical to exhaustively examine

every aspect of the area of interest, thus necessitating a compromise between the information obtained and the cost of obtaining it.

The cost of obtaining information will vary with the context. Where decisions must be made about large numbers of students in a short space of time, e.g. university admissions decisions, a premium is placed upon easily obtained information which can be processed quickly and efficiently. In such contexts aggregated numeric measures of academic performance are heavily relied upon. In other contexts, for example in selecting students in elementary school for learning-handicapped remedial programmes, extensive interview and testing would be undertaken to maximize the information required for this critical decision.

Tests are evaluation instruments designed to extract as much information about the performance criterion as can usefully be applied to the decision(s) to be taken, within the constraints imposed by the testing context, which incidentally will usually allow for only a single attempt by a student at the test.

The information extracted can be further increased if the test is tailored to fit the specific population of students concerned in the criterion decision; for example, question or item difficulties are chosen to spread scores (in a representative experimental sample

of students) over the full range of possible scores. For this and other reasons test development is an iterative process whereby test quality is enhanced by successive modifications to items, test length, and by repeated experimental administrations.

What then are the measures of test quality? Clearly, since the object of designing the test was to obtain information for some particular decision(s), the test scores obtained by target populations should correspond (in a way that is appropriate to the decision to be made) to the criterion of performance, i.e. the test score should be relevant to the decision. For example, if a group of students is to be ranked on the basis of mastery of some skill, then the test should (ideally) rank the students according to their degree of mastery, however that is independently established, i.e. by having an accepted concurrent criterion measure of mastery in the skill or by some other means.

Similarly if the test was designed to identify students in need of remedial help in some area, it should select the same people as the criterion selection would have done, e.g. panel of experts or teachers. Again if the test was designed to measure the gain in a student's learning after participating in a programme, the difference between (or ratio of) pre-test and post-test scores obtained for students in the programme should reflect the relative changes in level of learning as measured by other, more comprehensive means.

The degree of correspondence or relevance of test scores to the performance criterion is termed the EMPIRICAL VALIDITY of the test. If a quantified criterion of performance is available for the experimental group of students then validity can be quantified by appropriate measures of statistical association. The correlation coefficient of test scores with the criterion scores for a representative sample of the target population is termed the VALIDITY COEFFICIENT.

Since tests are administered in the main (to the target population at least) only once, it is desirable that the single score obtained be representative of the average score a student would obtain were the same or an equivalent test to be repeated a large number of times, (termed his "true" score) since uncertainty in the test score will lead to uncertainty as to the value of the criterion measure on which a decision is to be made. As the relative magnitude of the uncertainty (error) increases, so the utility of the test score decreases, particularly for decisions relating to differences between individual students. Clearly then observational "noise" results in a loss of information about a student's "true" score. This loss is indexed by the coefficient of RELIABILITY which may vary from a value

of one for tests which provide perfectly consistent scores for a given individual every time he/she takes the test, to a value of zero for a test where variations in an individual's score from test to retest are so large as to make comparisons between the scores of individuals meaningless.

Factors which can influence test reliability derive, by and large, from the following sources;

- inconsistency in the performance of individual test-takers;
- unreliability of the scoring process itself where the score assigned may vary with a rescoring. This source of reliability is greatly affected by the form of the test. Tests, such as essay tests, which are scored subjectively are prone to far greater variations between scorings of the same test than are multiple-choice tests which are objectively scored (in that no room is allowed for scorer subjectivity but mechanical errors may occur);
- variations in the ambient conditions of the test administration.

A number of remedies are available to test designers for improving the reliability of a test, for example, increasing the number of items in the test.

Another test quality consideration concerns the way in which the test scores of the target population (or a representative sample) are distributed over the range of possible scores. Even in the absence of observational "noise" a test measure will provide little information if it does not discriminate sufficiently between test-takers. In an extreme case a test on which the whole population achieves the same score contributes no information whatsoever.

Finally we turn to a quality of tests or examinations which affects the test-taker rather more than the institutional decision-maker. Since educational tests and examinations can have far reaching effects on the subsequent economic and occupational opportunities of the examinees, such instruments should be fair in the sense that the score achieved is not consistently biased by incidental characteristics of the test-taker, unrelated to the performance criterion. Such bias might arise from inappropriateness of test content for some groups in the target population, the influence of such incidentals as handwriting, or variations in institutional standards of evaluation. These biases all arise from subjectivity of either test-designers, standard-setters or scorers.

To sum up the desirable qualities of a test or examination measure, a good test should provide information which is relevant (VALID), precise (RELIABLE), efficient and fair, with respect to a target

population and a criterion decision. The context of testing will determine the primacy of one quality over another.

3.2.3 Achievement Testing in Ontario

Ontario had little need of standardized achievement testing for prospective university students until the mid-sixties. The conditions which precipitated the development of standardized admission measures in the U.S. had not applied, up to that time, to the Ontario educational system. Ontario universities served a homogeneously prepared clientele drawn, in the most part, from Ontario secondary schools, and evaluated on the basis of provincially set and administered examinations.

The sweeping educational reforms ushered in by the Hall-Dennis report Living and Learning - including opposition to centralized competitive examinations and uniform curricula - together with a swelling tide of students to be accommodated, led to the phasing out of departmental examinations by 1968.

By this time a national testing programme, the Service for Admission to College and University (SACU), had been incorporated and, with the support of the Ontario Department of Education (payment of test fees)

by 1972-73 almost all grade 13 students were writing the SACU tests. Lack of support for the programme by universities and the withdrawal of support by the Department of Education resulted in the demise of SACU as an operating organization at the end of 1974. It was felt by many in the universities that standardized aptitude tests of verbal and mathematical ability were not capable of replacing teachers' marks as predictors of university success and consequently the test scores were not being extensively used in the admissions process.

The underlying need for standardized measures of pre-university performance continued, however, with greater emphasis being given to achievement and placement tests rather than to the admissions screening of scholastic aptitude tests. Implicit in this changing mood was the determination of the universities to accommodate the diversity in preparation of new registrants by internal reorganization of instruction (alternative courses and course sections, remedial basic skills courses) rather than by selection.

3.3 CTEL and CMPT

3.3.1 COU Fall 1975 Testing Programme

In June 1975, the Council of Ontario Universities initiated an experimental programme of achievement testing by purchasing, for a nominal fee, two standardized tests, one of English and one of

mathematics, developed for the SACU battery, from the Association of Universities and Colleges of Canada. The tests were administered in September 1975 to freshmen registering at four Ontario universities: Brock, McMaster, Queen's and Waterloo. Participation was not compulsory and less than 60% of eligible students wrote a test.

Table 1 gives the breakdown of test-takers and reference populations by programme of registration, in order to illustrate the degree to which test-takers were representative of these populations with respect to programme of registration. A more detailed breakdown by programme and university will be found in the appendix, Table A1.

It can be seen from Table 1 that the CTEL sample was made up in much the same way, by programme of registration, as was the total 1975 COU registrant population. Neither group, however, is composed in the same way as is the cohort of first year registrants in all Ontario universities. This latter population contained a far larger proportion of arts students (42.3%) than did either total COU registrants (31.2%) or the CTEL volunteers (30.7%). As for CMPT, not unexpectedly, very few students who finally registered in an arts programme wrote the test. The testing design specified that only those students initially registering in programmes for which some year 5 mathematics was a requirement, were eligible to write CMPT. If only those students registering in

science, engineering or business programmes are considered, then the test and reference populations are made up of science, engineering and business students in the following proportions:

	Science	Engineering	Business
CMPT volunteers	50.9%	34.4%	14.6%
All COU '75 registrants	54.9%	30.8%	14.3%
All Ontario Universities	60.3%	21.7%	17.9%

The CMPT volunteer group has a higher proportion of engineering registrants (34.4%) and a lower proportion of science registrants (50.9%) than does either of the reference populations. The difference in composition is most pronounced between the CMPT volunteer group and the all-Ontario population.

Table 2 shows the breakdown of the tested groups and total freshmen at the four participating universities by broad categories of educational source. This classification is based upon the educational qualifications or background on which a student's application for admission to an Ontario university was assessed.

Although the total COU registrant group had much the same composition by educational source as did the population of first year registrants

Table 1

COU 1975: Programme of Registration for CTEL, CMPT Achievement Test Writers and all First Year Registrants at Brock, McMaster, Queen's and Waterloo, and for all Ontario Universities

Programme of Registration[a]	1975 COU Universities						All Ontario Universities[b]	
	CTEL		CMPT		All Registrants			
	No.	%	No.	%	No.	%	No.	%
Arts (Humanities & Social Sciences)	1304	31.4	30	1.2	3221	32.6	16106	45.9
Sciences (Natural Sciences & Mathematics)	1287	31.0	1149	47.8	2749	27.8	8259	23.6
Engineering & Applied Science	631	15.2	777	32.3	1541	15.6	2977	8.5
Environmental Studies	188	4.5	39	1.6	394	4.0	395	1.1
Business	302	7.3	330	13.7	713	7.2	2454	7.0
Physical & Health Education and Recreation	366	8.8	76	3.2	901	9.1	1641	4.7
Nursing and Rehabilitation Medicine etc.	46	1.1	1	-	179	1.8	777	2.2
Other	32	0.8	1	-	178	1.8	2448	7.0
Total with Identified Programme	4156	100.0	2403	100.0	9876	100.0	35057	100.0
Missing Programme Information	96		68		437		2985[c]	
TOTAL	4252		2471		10313		38042	

a. This was the latest programme of registration available for each university.
b. Source: Application Statistics 1975, Council of Ontario Universities, November 1978.
c. Includes registrants in first year post-year 5 programmes who did not apply through the Ontario Application Centre e.g. repeaters, internal transfers etc.

Table 2

COU 1975: Educational Source of all First Year Registrants and Achievement Test Volunteers at Brock, McMaster, Queen's and Waterloo, and of all Ontario universities

Educational Source[a]	1975 COU Universities						All Ontario Universities[b]	
	CTEL		CMPT		All Registrants			
	No.	%	No.	%	No.	%	No.	%
Ontario Year 5 (Current)	3535	83.8	2213	90.2	7427	78.8	26902	76.7
Ontario Year 5 (Prior Year & Preparatory)	231	5.5	77	3.1	610	6.5	2682	7.7
Year 5 Equivalent (Other Canadian Province)	43	1.0	44	1.8	199	2.1	700	2.0
Year 5 Equivalent (Other Country)	81	1.9	37	1.5	256	2.7	788	2.2
Transfers from Post-Secondary Sources	149	3.5	62	2.5	580	6.2	1949	5.6
Miscellaneous (Mainly Ontario Year 4 & Adult Students)	181	4.3	20	0.8	349	3.7	2036	5.8
Total	4220	100.0	2453	100.0	9421	100.0	35057	100.0
Repeaters, Internal Transfers & Unidentified Sources	32		18		892		2985	
Total	4252		2471		10313		38042	

a. Source: Based upon the Ontario Application Centres' data element 340 - Educational Source of Applicant

b. Source: Application Statistics 1975: Council of Ontario Universities, November 1978

in Ontario universities at large, the same cannot be said for the students who wrote CTEL and CMPT. In comparison to the two reference populations the CTEL and CMPT groups had higher proportions of current year 5 graduates, i.e. 83.1% of CTEL writers and 89.6% of CMPT writers compared with 72.0% of all COU registrants and 70.7% of all Ontario first year registrants. If only "new" entrants to first year from identifiable sources are considered, i.e. repeaters, internal transfers and unidentified sources are excluded from the calculations, then the proportions of current year 5 graduates in each group become 83.8% of CTEL test-takers, 90.2% of CMPT test-takers, 78.8% of all COU registrants and 76.7% of all Ontario first year registrants. Clearly the tested groups were somewhat more representative (by educational source) of the "new" entrants than they are of first year registrants as a whole.

As regards the representativeness of the test-takers with respect to high school academic performance measures (1974-75 year 5 graduates) the CTEL volunteer group had somewhat higher mean marks in most year 5 subjects than did all COU registrants. For year 5 English, biology and French the difference in mean marks between the CTEL group and that of all COU students was statistically significant. The magnitude of these differences in group means was less than one percentage point for all year 5 subjects. The students who volunteered to write CMPT showed evidence of higher achievement in year 5 subjects than either the CTEL group or the complete COU cohort.

The differences in group mean marks were most pronounced for mathematics and science subjects. Not all of this difference can be accounted for by the fact that only registrants in progammes with some year 5 mathematics requirements were eligible to take CMPT. As Tables 13.1 and 13.3 show (see pages 82 and 84) the year 5 averages of CMPT volunteers within each programme were almost always higher than year 5 averages for all current year 5 graduates in that programme. The difference in means was small however; of the order of one percentage point. No data on year 5 mean subject marks for all first year entrants from year 5 to Ontario universities in 1975 were available for comparison with the 1975 COU groups.

A regression analysis of CTEL and CMPT scores for the volunteer groups on these tests, using year 5 averages as predictors, was employed to estimate the average test scores which all freshmen from year 5 would have obtained had they written tests. The results of this adjustment indicate that the mean CTEL total score obtained by the volunteer sample (40.3) should be deflated somewhat to about 39.9 for all year 5 graduates entering the four universities.

Very little concerning aggregate performance variation can be inferred from the two previous experimental administrations of CTEL, the first being a national sample of grade 12 and 13 students drawn from all provinces (several years ago) and the second being administered to one faculty of engineering (University of Toronto) and two English departments in three Canadian universities - one in Western Canada, one in Central Canada, and one in the Atlantic Provinces - in 1974.

A comparison of the mean CTEL total raw scores obtained in the three administrations are presented below:

	Number in sample	Mean	Standard Deviation	Standard Error of mean
National Sample	1,876	41.9	13.1	0.30
1974 Sample	3,308	38.7	11.9	0.21
1975 COU (CTEL volunteers)	4,546	40.1	12.9	0.20

Although the differences in mean level of performance among these samples are statistically significant, any further interpretation would be speculative given the compositional differences between the groups tested.

The conclusion of the foregoing brief discussion of comparisons between aggregate levels of performance on CTEL, is that no inferences can be made with respect to changes in mean performance of any replicable reference population which might prove useful in determining a longitudinal trend. The mean test scores broken down by programme within the participating universities in the COU study could be used in future years by these same institutions to monitor trends, should they decide to administer CTEL. Other Ontario universities and/or faculties intending to use CTEL could relate their own estimated freshmen levels to the experimental groups by equating mean test scores through mean high school averages, or better still by using the regressions of CTEL on high school average for equivalent programmes.

Reference levels are not available for CMPT simply because no previous large scale administrations have been attempted. The dearth of comparative

findings is not serious however; it has little impact on the primary research questions and test development objectives of the COU project.

3.3.2 Technical Quality of CTEL and CMPT

In contrast to the lack of useful comparative data on CTEL and CMPT, a wealth of technical data has now accumulated. Tables 3 and 4 present some of the internal characteristics of CTEL and CMPT respectively, as estimated from the 1975 COU administrations.

Coefficients of reliability for CTEL subtest scores are all in the region of 0.8 while that of the total score is 0.84. The precision of measurement afforded by each CTEL subtest and total score is more readily interpreted by looking at the standard error of measurement column of Table 3. According to the classical model of test theory, if the test were to be administered to the same examinee a large number of times, the resulting scores would vary about some average value (termed the examinee's "true score") with a standard deviation equal to the standard error of measurement. Given an examinee's test score and the standard error of measurement for the test, an interval can be established, centred on the observed test score, which will contain the "true score" at some prescribed level of probability. For example, should a student obtain a total raw score of 40.0 on CTEL, we would expect his/her "true score" to lie between 34.2 and 45.81 about 68% of the time and between 28.6 and 51.4 about 95% of the time.

Table 3

CTEL: Internal Characteristics - Total Sample

Section	No. of Items	Time Allowed (Mins)	RAW SCORE Mean	RAW SCORE S.D.	Correlations between section scores 1	2	3	Total	KR-20 Reliability	S.E. of Measurement
1 Construction Shift	22	20	10.99	4.41		.61	.69	.86	.80	1.98
2 Prose Groups	28	20	14.09	5.26			.63	.87	.81	2.32
3 Usage	30	20	15.03	5.08				.89	.79	2.35
Total	80	60	40.11	12.89					.84	5.81

N=4546

Table 4

CMPT: Internal Characteristics

Section	No. of Items	Time Allowed (mins)	RAW SCORE N	RAW SCORE Mean	RAW SCORE S.D.	Correlations between section scores A1	A2	B1	B2	B3	α Reliability	S.E. of Measurement
A1 Algebra	25	30	1266	15.61	5.20			.79	.73	.59	.85	2.01
A2 Calculus	25	30	1349	11.21	4.25			.70	.64	.64	.78	1.99
B1 Trigonometry & Functions	25	30	2352	14.66	5.51	1041	1221		.75	.69	.87	1.99
B2 Synthetic & Coordinate Geometry	25	30	1856	15.09	4.76	885	892	1540		.62	.82	2.01
B3 Matrices, Vectors & Trans Geometry	20	30	997	7.69	4.00	451	498	722	243		.78	1.88

Entries below diagonal are numbers of students writing the two sections indicated. These are the N's on which correlation coefficients above the diagonal are based.

Volunteers were requested to write

a) one of A1 or A2
and b) two of B1, B2 or B3.

Cross-correlations among CTEL and CMPT subtests range from 0.27 to 0.36, and, together with the high inter-correlations within the CTEL and CMPT batteries as shown in Tables 3 and 4 indicate that the two tests form mutually distinguishable and internally consistent sets of measures.

It should be pointed out that these tables merely summarize a few of the important internal characteristics of the two tests. Details of item difficulties, item discrimination power, speed and raw score frequency distributions are contained in the technical reports.

CTEL subtest and total test reliabilities are respectably high considering the widely differing difficulty levels of items and the short length of each section. The same can be said of the reliability data on CMPT although some improvements have been suggested[5]. Although CMPT was designed as a placement test, some institutions might wish to use it as a general test of mathematics achievement for limited enrolment programmes especially for non-year 5 graduates. The provision of an equating formula to estimate a total score from subtest scores (assuming not all subtests were required) would be useful in such cases.

---

[5]Vincent D'Oyley et al, The Canadian Mathematics Placement Test (CMPT) Technical Report, August 1976

The two tests, then, are of good quality from a purely technical stand point, as precise measuring instruments for their intended population. Findings as to what they might possibly be measuring are dealt with next.

3.3.3 Validity Evidence for CTEL and CMPT

In the discussion of 3.1 it was argued that the validity of a test was properly defined in terms of the relevance of scores obtained, by members of the target population, to some educational decision. Such validity may be empirically demonstrated in a sample from the target population by the strength of the association of sample test scores to corresponding "scores" on criterion measures, where these are available. The criterion measures are those that would have been used in making the particular decision were they available at the time when the decision was to be made, or were it economical to ascertain their values. The test score will substitute for the criterion measure in making the decision. Thus, for admissions decisions, where an attempt is being made to select students according to their future academic performance in the institution or programme (performance which clearly cannot be measured at the time of the admissions decision) test scores of applicants should be distributed in a way similar to their eventual marks. The correlation coefficient between criterion and predictor scores for an

experimental sample summarizes the strength of the relationship which is being sought. In this instance, the correlation coefficient would quantify the PREDICTIVE VALIDITY of the test.

In other circumstances a test might be used as a proxy for a measure of current academic performance which is either unavailable or too costly to obtain. Two topical examples would be:

- the testing of applicants who do not come from the Ontario secondary school system, to determine their equivalent achievement levels;

- testing to determine adequacy of writing skills where no reliable prior data exist and where comprehensive evaluation is impractical.

Again, correlation coefficients between test scores and corresponding criterion scores on the current measure, obtained from a representative sample of the population about which the decisions are to be made, estimate the test's validity, this time being termed CONCURRENT VALIDITY.

Another type of evidence of a measure's validity of considerable importance in the evaluation of achievement tests may be obtained by examining test-items and verifying that they are representative of subject matter or skills required by the test objectives. This type of validity is known as CONTENT VALIDITY.

It may be claimed by test designers, inferred from test content, or discovered fortuitously that performance on the test reflects the degree to which the test-taker posesses some hypothetical trait or ability (termed a CONSTRUCT). The construct is considered to affect measured performance in a variety of situations e.g. a construct such as mathematical aptitude would be expected to influence positively year 5 marks in mathematics and physics, but not in English. The extent to which test scores are found to relate to other available measures in a way that is consistent with the expected effect of the construct is considered as evidence for CONSTRUCT VALIDITY. Validation evidence of this latter kind is often introduced to study traits or abilities for which no adequate criterion measure can be found. One such construct, of particular interest in the present study, might be labelled "standard English prose writing ability appropriate to students entering Ontario universities". As noted later, neither year 5 English marks nor first year university English marks can be considered as satisfactory criterion measures of this construct. However, some idea of the capacity of tests such as CTEL and RCLAT to measure writing ability can be obtained by noting the degree of association between the test scores and marks awarded in subjects for which writing skills are expected to be evaluated to a greater or lesser degree.

It is customary when reporting the results of experimental test administrations to present some correlations of test scores with promising candidates for criterion measures of predictive or con-

current validity without reference to their interpretation for any specific decision, e.g. correlations of CTEL scores with year 5 English marks and with first year university English marks. It is also customary to include correlations of test scores with other measures which have no theoretical pretensions as criteria for the test. All else being equal, higher correlation coefficients of test scores with the designated criterion measures, than correlations with theoretically unrelated measures, are considered as evidence of DISCRIMINANT VALIDITY (a type of construct validity).

Tables 5 and 6 show some test-criterion correlations for CTEL and CMPT respectively. Suitable machine readable student records for Brock were not available, so that only McMaster, Queen's and Waterloo results are shown for university performance.

The year 5 and university first year criterion measures shown are just a few of the many that might have been selected. They have been chosen to give some idea of the concurrent, predictive and discriminant validities of each test. Correlations with university criterion measures are also shown for year 5 mathematics and English averages so that their predictive validities may be compared with those of the tests.

When attempting to interpret such tables a number of factors which affect the size of correlation coefficients between academic performance measures, quite apart from the validity relationships between them, must be taken into consideration.

a) Correlations between unreliable measures will underestimate the strength of the statistical relationship between "noise-free" measures of the same performance. The reliabilities of measures formed by aggregating over several measures similar in nature, will in general, show an increase over the reliability of a single measure. Reliabilities may also be affected by the marking policies and record-keeping procedures of the institutions supplying the data upon which correlations are computed. It must be appreciated that the primary purpose of university student record-keeping is not to facilitate predictive validity studies, and therefore some student records will contain spurious data, not representing true academic performance. "Noise" introduced in this way will reduce reliabilities and validity coefficients. Since this effect varies from institution to institution direct comparisons of validities between universities is hazardous.

b) Because university students are <u>selected</u> from the general pool of applicants on the basis of high school marks, the range of their scores on this measure is relatively restricted. This reduction in range tends to produce lower predictive validity coefficients than would be observed in an unselected group. The severity of the effect will vary over universities (and programmes) in accordance with the degree of selectivity operating in their admissions to first year. The measures most

affected will be those upon which direct selection has been made, year 5 averages for example, rather than test scores.

As a criterion measure of English language achievement upon completion of Ontario year 5 studies, the year 5 English average is somewhat heterogeneous in nature, since course curricula vary widely in content from school to school. Ministry of Education course codes do not allow the isolation of year 5 English courses concerned primarily with the development of English language skills, and so we are forced to accept the undifferentiated measure as the best available concurrent (with CTEL) measure of year 5 English language achievement. Despite these misgivings, Table 5 shows clearly that all of the CTEL subtests relate more strongly with year 5 English than with other concurrent measures, and moreover, the ordering of relationships appears to be consistent with intuitive expectations, e.g. year 5 history achievement should be closer to English achievement than are mathematics and chemistry. Thus CTEL's discriminant validity for end of year 5 English achievement is confirmed.

Table 3 showed that estimated reliabilities for the three subtests were very nearly the same, so that the higher correlations of CTEL - 3 (usage items) with year 5 English and year 5 history suggests a higher concurrent validity for this section .

Turning to predictive relationships between CTEL scores and first

Table 5

Correlations of CTEL Scores with High School and University Subject* Averages

CTEL		Year 5 Subject Averages (1974-75)					University Final Year 1 Subject Averages (1975-76)					
							ENGLISH			MATHEMATICS		
		English	History	French	Math.	Chemistry	McMaster	Queen's	Waterloo	McMaster	Queen's**	Waterloo
CTEL-1	Construction Shift Items	.38	.36	.31	.27	.32	.27	.33	.35	.12	-	.28
CTEL-2	Prose Group Items	.42	.34	.29	.25	.30	.23	.39	.28	.11	-	.30
CTEL-3	Usage Items	.46	.40	.38	.24	.31	.36	.35	.31	.11	-	.26
CTEL	Total	.49	.42	.37	.29	.35	.33	.41	.37	.13	-	.32
	N	2906	1130	907	3037	2451	333	117	314	878	-	1041
YEAR 5 ENGLISH AVE.							.47 [468]	.54 [364]	.40 [614]	.26 [983]	.28 [764]	.34 [1202]
YEAR 5 MATH AVE.							.26 [1255]	.28 [229]	.27 [447]	.40 [1255]	.53 [958]	.57 [1759]

* Course marks are credit weighted to produce subject average

** Only 12 students taking a mathematics course at Queen's also sat the CTEL

[ ] Figures in brackets are the numbers of cases upon which the adjacent correlation coefficient has been calculated

year university final marks in English and mathematics, Table 5 shows a not unexpected weakening of the degree of statistical association although CTEL subtests and total continue to discriminate between English marks and mathematics marks within each university. (Queen's mathematics correlations were not statistically significant, since only a few CTEL volunteers subsequently took a mathematics course.)

For purposes of comparison of relative predictive power, year 5 English and mathematics averages are shown correlated with first year university averages in the same subjects. Within each participating university the year 5 English average is more highly correlated with university English than is any CTEL measure. This does not, of course, imply that year 5 English average measures the "standard English prose writing" construct any better than CTEL does. It does however imply that a year 5 English mark gives a better prediction of first year university performance in English courses than does a CTEL score.

Some validity data for CMPT are presented in Table 6. Concurrent validities with year 5 mathematics course marks and averages are moderately high for all subtests, with A1-Algebra and B1-Trigonometry and Functions showing consistently higher values than the others. These two subtests have the highest estimated reliabilities (from Table 4) which partially explains their higher correlations with other measures. Since large numbers of students have attempted at least two credits in mathematics in their first year (as opposed to the few taking more than

Table 6

Correlations of CMPT scores with High School and University Course Subject* Averages

CMPT Sections	Year 5					University Final Yr. 1 Mathematics Average				
	Course Marks		Subject Average			McMaster		Queen's	Waterloo	
	Functions & Relations	Calculus	Algebra	Math.	English	a.	b.	a.	a.	b.
A1 - Algebra	.54 [911]	.53 [948]	.52 [867]	.58 [999]	.23 [734]	.41 [348]	.55 [195]	.37 [67]	.48 [454]	.52 [398]
A2 - Calculus	.46 [1115]	.47 [1209]	.49 [851]	.49 [1212]	.18 [898]	.28 [425]	.42 [223]	.60 [134]	.38 [508]	.46 [381]
B1 - Trigonometry & Functions	.52 [1852]	.51 [1956]	.54 [1529]	.57 [2006]	.22 [1481]	.34 [696]	.50 [377]	.43 [183]	.48 [885]	.56 [720]
B2 - Synthetic and Coordinate Geometry	.45 [1408]	.44 [1492]	.44 [1109]	.49 [1529]	.15 [1133]	.22 [529]	.37 [284]	.52 [118]	.38 [675]	.46 [524]
B3 - Matrices, Vectors & Transformational Geometry	.41 [781]	.41 [853]	.41 [790]	.46 [876]	.23 [640]	.28 [313]	.37 [170]	.49 [99]	.41 [363]	.47 [313]
YEAR 5 MATH AVERAGE						.40 [1255]	.67 [548]	.53 [958]	.57 [1759]	.61 [1284]
YEAR 5 ENGLISH AVERAGE						.26 [983]	.40 [346]	.28 [764]	.34 [1202]	.32 [824]

Notes 1. a) For all students taking at least one university mathematics course.
b) For only those students attempting two credits or more in university mathematics.

2. A figure in parentheses is the number of cases upon which the adjacent correlation coefficient is based.

one credit in English) we can improve the reliability of the criterion measure by computing correlations of test scores with first year mathematics average for students with at least two credits in that subject.

Clearly, year 5 mathematics performance is a good predictor of mathematics performance at the end of first year university. No doubt estimated predictive validities would be even greater were it not for the restriction in the range of this measure caused by its use in selection for more competitive universities and programmes. Estimated predictive validities for CMPT are less affected by such direct selection. CMPT subtests, A1-Algebra and B1-Trigonometry and Functions, show predictive validities ranging from 0.50 to 0.56, when the criterion measure is taken as the average of at least two credits in university mathematics.

3.4 RCLAT and the Writing Test

3.4.1 COU Fall 1976 Testing Programme

During registration week in September 1976, two forms of an objective test of Reading Comprehension and Language Achievement (RCLAT) were administered to a large proportion of the cohort of registrants (including repeaters) in first year programmes at Carleton University, Erindale College and McMaster University.

In addition, a randomly chosen sample of the McMaster students, 741 in all, were assigned to do the Writing Test consisting of a 250 - 350 word expository mode essay to be completed in 1½ hours. Each completed essay was scored 'holistically' on an integer scale of 1 through 10, by three markers independently.

Both the RCLAT and Writing Test had seen service in the Interface Study but had not been administered to students taking part in the university records survey, so that the link between test performance and subsequent university performance had not been made.

Table 7 provides a breakdown of test-takers and other registrants by programme. This table shows that in contrast to the 1975 COU test administrations, a high proportion (90%) of eligible students at Carleton and McMaster wrote the RCLAT.

A small number (21) of RCLAT records could not be matched with records from any other source and are presumed to have been mistakenly written by ineligible students. The proportional representation of the RCLAT group by programme of registration closely resembles that of all registrants at Carleton and McMaster. A somewhat lower proportion (39.2%) of Arts programme students (all at McMaster) wrote the essay test than were present in the whole Carleton, McMaster freshmen group (46.0%). This however reflects differences in programme make-up between the two universities. As far as McMaster was concerned, the

Table 7

COU 1976: Programme of Registration for RCLAT and Essay Test Writers; All First Year Registrants at Carleton and McMaster, and all First Year Registrants at Ontario Universities

Programmes of Registration[a]	1976 COU Universities						All Ontario Universities[b]	
	RCLAT		Essay		All Registrants			
	No.	%	No.	%	No.	%	No.	%
Arts (Humanities & Social Sciences)	2300	46.9	286	39.2	2572	47.4	15337	44.3
Science	893	18.2	171	23.4	981	18.1	8372	24.2
Engineering & Applied Science	637	13.0	86	11.8	669	12.3	2820	8.1
Business	586	11.9	89	12.2	652	12.0	2861	8.3
Journalism	168	3.4	-	-	168	3.1	192	0.6
Physical & Health Education	211	4.3	57	7.8	242	4.5	1636	4.7
Nursing and Rehabilitation Medicine	63	1.3	23	3.2	70	1.3	733	2.1
Other	49	1.0	8	1.1	70	1.3	2652	7.7
Total with Identified Programme	4907	100.0	720	100.0	5424	100.0	34603	100.0
Missing Programme Information	163		10		164		5433[c]	
TOTAL	5070		730		5588		40036	

a. This was the latest programme of registration available for each university.
b. Source: Application Statistics 1976, Council of Ontario Universities, November 1978.
c. Includes registrants in first year post-year 5 programmes who did not apply through the Ontario Application Centre e.g. repeaters, internal transfers etc.

essay writers were a representative group by programme of registration (see Appendix Table A2 for a finer breakdown of the test samples by programme of registration and university). Direct comparisons with the pattern of programme participation for all Ontario universities is made difficult by the large proportion with missing programme information for the all-Ontario group.

Table 8 suggests that the composition, by educational source, of RCLAT writers was similar to that of all first year registrants at the participating universities and not significantly different from that of first year registrants at all Ontario universities.

For McMaster, 730 essay scores have been matched with university records, thus representing 24% of first year registrants in the fall of 1976. The pattern of representation of essay writers by programme of registration and by educational source is similar to that of the RCLAT writers, i.e. the Social Science I programme is slightly under-represented and Ontario year 5 current entrants are somewhat over-represented.

Unfortunately, although test data (RCLAT) have been analyzed and reported for the Erindale College test-writers, their subsequent first year university records were not available at the time of analysis and no further information can be provided for this group.

Table 8

COU 1976: Educational Source of RCLAT and Essay Test Writers, All First Year Registrants at Carleton and McMaster, and All First Year Registrants at Ontario Universities

Educational[a] Source	1976 COU Universities						All Ontario Universities[b]	
	RCLAT		Essay		All Registrants			
	No.	%	No.	%	No.	%	No.	%
Ontario Year 5 (Current)	3694	83.6	637	91.7	3914	80.0	27,195	78.6
Ontario Year 5 (Prior Year & Preparatory)	305	6.9	30	4.3	351	7.2	2,482	7.2
Year 5 Equivalent (Other Canadian Province)	74	1.7	3	0.4	74	1.5	714	2.1
Year 5 Equivalent (Other Country)	41	0.9	4	0.6	52	1.1	679	2.0
Transfers from Post-Secondary Sources)	257	5.8	13	1.9	443	9.1	2396	6.9
Miscellaneous (Mainly Ontario Year 4 and Adult Students)	48	1.1	8	1.2	59	1.2	1137	3.3
Total	4419	100.0	695	100.0	4893	100.0	34603	100.0
Repeaters, Internal Transfers & Unidentified Sources	652		35		695		5433[c]	
TOTAL	5071		730		5588		40036	

a. Source: Based upon the Ontario Application Centre's data element 340 Educational Source of Applicant

b. Source: Application Statistics 1976: Council of Ontario Universities, November 1978

c. Source: Includes registrants in first year post-year 5 programmes who did not apply through the Ontario Universities Application Centre e.g. repeaters, internal transfers etc.

3.4.2 Technical Quality of RCLAT and the Writing Test

Table 9 summarizes some of the major results on the internal characteristics of RCLAT and the Writing Test. Estimated coefficients of reliability for RCLAT subtests range from 0.53 to 0.70 according to which subtest, test form and university is specified. For the total test, reliability coefficients were slightly over 0.8 on Form 1 and slightly lower than 0.8 on Form 2.

Coefficients of correlation among scores on the three test parts are in the region of 0.5. When a correction is made for the attenuation in these coefficients caused by the unreliability of the subtest scores, inter-part correlations are found to range between 0.74 and 0.90 with a median value of 0.78. This compares to a range of corrected inter-part correlations for CTEL of 0.76 to 0.87 (median 0.79) and for CMPT of 0.72 to 0.91 (median 0.84). The square of the correlation coefficient between two sets of scores can be interpreted as the proportion of variance in one set which can be predicted from the variation in the other set. Thus using the median values for each test, approximately 60% of the true score variance in any set of CTEL or RCLAT subtest scores can be predicted from the corresponding true scores on another subtest. For CMPT the comparable figure is 70%. We may conclude that

Table 9

Internal Characteristics of the Reading Comprehension and Language Achievement Test (English) and the Writing Test

University	No. of items	Time Allowed (mins)	Correlations among RCLAT scores on Subtests and Total Test				FORM 1 Raw Score Mean	FORM 1 Raw Score S.D.	FORM 1 Reliab. Coeff.	FORM 2 Raw Score Mean	FORM 2 Raw Score S.D.	FORM 2 Reliab. Coeff.
1. Carleton												
Test Part			1	2	3	4						
1. Reading Comprehension Subtest	10	15	-	.48	.45	.76	5.44	2.64	.68	4.45	2.54	.59
2. First Language Achievement Subtest	11	10	.51	-	.56	.83	4.96	3.04	.69	4.49	2.79	.70
3. Second Language Achievement Subtest	15	15	.49	.53	-	.85	8.59	3.21	.61	6.72	3.26	.58
4. Total Test	36	40	.79	.84	.84	-	19.00	7.32	.83	15.66	7.03	.80
							N = 1272			N = 1270		
2. Erindale												
Test Part			1	2	3	4						
1. Reading Comprehension Subtest	10	15	-	.44	.49	.78	4.48	2.73	.69	3.48	2.44	.56
2. First Language Achievement Subtest	11	10	.49	-	.44	.77	3.61	2.79	.62	3.51	2.58	.63
3. Second Language Achievement Subtest	15	15	.57	.50	-	.84	6.95	3.42	.63	5.80	3.16	.53
4. Total Test	36	40	.82	.79	.86	-	15.05	7.38	.82	12.79	6.54	.77
							N = 555			N = 557		
3. McMaster												
Test Part			1	2	3	4						
1. Reading Comprehension Subtest	10	15	-	.48	.45	.76	5.08	2.62	.66	4.03	2.48	.55
2. First Language Achievement Subtest	11	10	.50	-	.53	.83	4.30	2.98	.67	4.04	2.80	.69
3. Second Language Achievement Subtest	15	15	.48	.53	-	.84	7.97	3.12	.57	6.31	3.25	.57
4. Total Test	36	40	.78	.83	.83	-	17.35	7.13	.81	14.38	6.94	.79
							N = 1274			N = 1276		
WRITING TEST	250-350 Word Essay	75	.42	.39	.45	.51	6.4	1.48	.76*			
							N = 730					

Note: Form 1 correlations are below the diagonal, Form 2 correlations are above the diagonal.

*This is the estimated reliability coefficent for the scoring process only.

compared to the CMPT battery, CTEL and RCLAT taken as a whole are relatively heterogeneous in content.

In comparing reliabilities for RCLAT with those estimated for CTEL (Table 3), the fact that the former test is much shorter - 40 mins. and 36 items compared to 60 mins. and 80 items for CTEL - should be taken into account. Even so, the reliabilities of the two total tests are not so very different, 0.8 for RCLAT compared with 0.84 for CTEL. However, inferences about individual differences on RCLAT subtest scores would be less precise than those based on CTEL subtest scores.

There are clearly good reasons for lengthening RCLAT to improve reliability if classification decisions about individuals are to be made on the basis of a single score. However, an alternative testing strategy in which Form 1 was administered to all students first and those scoring significantly higher than the cut-off score eliminated from further testing (with Form 2) would reduce overall testing time considerably while maintaining the same level of statistical precision.

For the Writing Test, the reliability of the scoring process was estimated to be 0.76 and the standard error of measurement 0.90. These figures do not take into account the likely variability in a student's own performance over repeated writing tasks. Thus no realistic lower bound to the reliability of the Writing Test as a whole has been estimated. With reliability levels as low as 0.76, quite large

differences between the scores of two students would be needed before it could be confidently declared that the higher scoring student was the better writer. This fact however, does not rule out the use of the Writing Test for group comparisons and as a coarse screening device.

The results of technical analyses of RCLAT and the Writing Test for the COU administration are in close agreement with those reported in the Interface Study,[6] even though there were some differences in test administration procedures, (Interface students benefitted from study of a student handbook describing the tests and giving practice questions) and in the time of the academic year at which testing took place, (Interface students were tested near the end of the high school year whereas COU students were tested during registration week in the university fall term). The foregoing differences in test procedures between the COU and Interface English test administrations, could well have contributed to lower aggregate test performances by the COU groups. Furthermore, the scores of the Erindale RCLAT writers in the COU administration were probably affected to some degree by the Metro Toronto teachers' strike. The distribution of RCLAT total scores for the 2255 students in the Interface SSHGD group (a sample of Ontario year 5 Anglophone students in 1975-76) had a mean of 17.5 and a standard deviation of 7.0 after all scores had been converted to the form 1 scale; on the Writing Test, 569 SSHGD students obtained a mean score of 6.4 with a standard deviation of 1.3. The corresponding results for the

---

[6] R. Traub, et al, Secondary-Post-secondary Interface Project II: Nature of Students, Vol. 2, 1976, pp. 113-115, and pp. 440-441

COU administration were: a mean of 17.03 and a standard deviation of 7.12 for the RCLAT total score (N=6204), and a mean of 6.4 and standard deviation of 1.48 for the Writing Test (N=730).

Scores on the Writing Test were correlated more highly with RCLAT subtest and total scores in the COU (McMaster) sample (.51) than in the Interface sample (.46).

As regards the possibility of substituting objective tests of English language for writing samples, the question naturally arises as to what might have been the correlation between a comprehensive and perfectly reliable measure of writing ability and a perfectly reliable objective test score. On the basis of the results obtained in the present study we can only report that a single short essay scored with perfect reliability would correlate 0.64 with a perfectly reliable RCLAT total score.

3.4.3 Validity data for the RCLAT and Writing Test[7]

Table 10 shows correlations of RCLAT subtests scores, total score and calibrated essay scores with some concurrent achievement measures - year 5 averages - and some possible measures for predictive validity taken from end of session final marks at two universities, Carleton and McMaster.

---

[7] Potential validity data for RCLAT and the Writing Test were obtained in the Interface Study Project II. Ibid., pp. 440-441

Year 5 Direct
(1975-76) only

Table 10

Correlation Coefficients of RCLAT and Writing Test Scores with other Academic Performance Measures

		Year 5 Subject Averages (1975-76)										University Final Year 1 Subject Averages (1976-77)					
		English		History		French		Math.		Science		English		History		Math.	
	RCLAT	Carl.	McM.	Carl.	McM.	Carl.	McM.	Carl.	McM.	Carl.	McM.	Carl.	McM.	Carl.	McM.	Carl.	McM.
1.	Reading Comprehension	.33	.38	.32	.30	.22	.21	.21	.21	.26	.30	.30	.31	.25	.28	.27	.14
2.	First Language Achievement Test (construction shift items)	.38	.34	.27	.29	.27	.27	.26	.25	.29	.26	.31	.25	.14	.21	.26	.16
3.	Second Language Achievement Test (sentence correction items)	.38	.36	.24	.27	.30	.26	.21	.22	.24	.22	.31	.28	.14	.22	.21	.13
	RCLAT TOTAL	.46	.45	.36	.35	.33	.30	.28	.28	.32	.31	.38	.34	.21	.29	.28	.17
	N	1411	1868	558	560	605	522	1102	1813	1194	1770	472	455	351	522	703	1248
	Writing Test		.50		.30		.34		.12		.16		.40		.42		.14
	N		559		153		170		539		517		151		108		364
	YEAR 5 ENGLISH AVERAGE											.57 [462]	.38 [492]	.46 [340]	.45 [411]	.35 [587]	.26 [1110]
	YEAR 5 MATH AVERAGE											.53 [228]	.36 [323]	.54 [160]	.36 [277]	.68 [685]	.42* [1334]

* For 544 students with 2 or more attempted credits in mathematics the coefficient rises to .67.

Figures in parentheses show the number of cases upon which the adjacent coefficient has been based.

As was the case for CTEL, the RCLAT scores correlate more highly with year 5 English average than with other concurrent measures. Again, the relative ordering of the degree of relationship is consistent with the hypothesis that RCLAT subtests and total, measure English language achievement more than some other trait such as quantitative ability. This discriminant property is even more evident for the Writing Test which correlates 0.5 with year 5 English but only 0.12 with year 5 mathematics average.

The predictive validities of the RCLAT total score with first year final marks as criteria are similar to those of CTEL total (for the only directly comparable university, McMaster.) Table 10 shows Writing Test scores to be correlated somewhat more highly with year 5 and first year university English averages than are RCLAT scores[8]. The Writing Test also discriminates more sharply between humanities and mathematics/science performance, at the year 5 level particularly. Bearing in mind the lower reliability of the Writing Test as compared with the RCLAT total score (which causes a more pronounced attenuation in its correlation with any other measure) it appears that a Writing Test is potentially a more valid measure of the "writing skills" construct than the objective test. Its practical effectiveness, however,

---

[8] The Writing Test correlations are based upon the performance of a subset of the students who wrote RCLAT and hence their comparability with the RCLAT results may be questioned. In fact, the statement made above applies equally well to students taking both tests.

is limited by the unreliability of the scoring process to about the same level as that of the RCLAT total score. A composite measure, combining RCLAT subtest scores with the Writing Test score raised concurrent validity with year 5 English to 0.60 (N=207), and predictive validity with McMaster first year English average to 0.47 (N=150).

When comparing the predictive validities of the year 5 English and mathematics averages, it should be taken into account that the latter measure is based upon about twice as many marks, on average, as is the former and hence should have a higher reliability.

Only a cursory examination of the validity evidence for the achievement tests administered under the auspices of COU, has been presented thus far. The evidence which has been presented can only be described as a kind of informal construct validity, pointing out the strong and weak statistical relationships between the academic performance measures of interest. Subsequent analyses will attempt to supply validity data for specific test uses.

4.0 ANALYSES

4.1 Introduction to Analyses

The preceding chapter dealt with measures of academic performance which might be used by universities in admissions and placement decisions for new entrants. The discussion focussed on the technical quality of standardized tests - two in English and one in mathematics - which were administered experimentally in 1975 and 1976 for purposes of test development and the evaluation of test utility in decision-making.

The analyses presented in the present chapter are an attempt to determine what can be inferred from the COU experimental test administrations concerning the utility of the tests in question for a number of student classification decisions for which traditional data are currently seen as inadequate or inappropriate. Standardized achievement tests are also potentially useful in monitoring changes in aggregate group achievement levels for purposes of diagnosing strengths and weaknesses in preparation in specific subject areas with implications for modifying pre-university programmes, or for planning to accommodate to such changes within universities.

The experimental administrations by COU in 1975 and 1976 have added very little that is useful to knowledge in the latter area. The participating universities might use the aggregate results as a

benchmark were they to administer exactly the same tests under the same conditions to future first year registrants. Other universities would have less use for these aggregate results unless they were prepared to adopt one or other of the participating universities or programmes as a standard for their own students (before they have accumulated test data for themselves). Also since the tests were administered with a view to their development and refinement, a number of modifications have been suggested as a result of the experimental administrations.[9] Such revisions as lengthening a test or increasing its administration time will almost certainly lead to a change in the score distribution when the revised test is administered to a group of students having abilities comparable to those in the initial administration. However, it should prove possible to equate scores on the COU forms with those on corresponding revised versions by administering both tests to a representative group at some future date. The benchmark established in the COU administrations could then be adjusted to reflect equivalent performance levels on the revised tests.

---

[9] For example, an increase in the time allowed on the subtest CTEL 1 - "Construction Shift" has been suggested to counteract its apparent speededness. For CMPT it may be necessary to add more items and include some more intricate problems on the Calculus subtest and increase administration time on all subtests but Al-Algebra. RCLAT should be lengthened to improve its reliability, possibly by having students write both forms of the test. Otherwise an increase in administration time (for the first two parts in particular) is called for.

Needless to say, no inferences as to trends in aggregate standards of first year entrants to Ontario universities as a whole or even for those entering the participating universities, are warranted from the data analysed here. Two observations of the same population with the same evaluation instrument are required to determine a trend.

The foregoing arguments apply equally well to any uses of the COU data for establishing "norms" of performance against which decisions about individual students are to be made. However, the kinds of modifications which are likely to be made to the tests examined here should not unduly affect findings concerning test utility for classification decisions. Changes such as lengthening a test or increasing administration time are aimed primarily at improved reliability. The results, based upon the experimental sample would tend to underestimate the utility possible with modified tests.

Apart from the monitoring of aggregate standards, all other uses of standardized (and unstandardized) measures of student academic performance relate to some form of classification of individual students. Recapitulating the discussion of issues given in Section 2, the situation in which Ontario universities presently find themselves is imposing a greater burden of classification on admisssions personnel and teaching faculty, both in the range and sophistication of decisions to be made. No longer is the uniformity of curricular preparation and

of evaluation measures, which once characterized the Ontario grade 13 - first year university interface (both attributable in large part to the anchoring influence of "departmental examinations") a fact of life. Nor can universities rely upon the high selection ratios (applicant to registrant) - which once ensued from the supply of university places lagging behind the demand for admission - to limit the range of variation in the ability levels of registrants. Add to this a growing group of applicants from previously untapped sources such as mature students and those from non-Ontario jurisdictions, and it is evident that Ontario universities must deal with an unprecedented degree of diversity in the abilities and academic backgrounds of incoming students.

Increased selectivity in general admissions is no longer an avenue which commends itself to serious consideration as a means of narrowing the range of abilities among admitted students. The selection measure likely to be used (school marks) is itself in doubt, biased by school and probably adaptive to changes in minimum entrance standards; increased selectivity would involve considerable shrinkage in the system as a whole which is unlikely to be appreciated by parents of rejected students or by displaced faculty; and it is a wasteful way of increasing the quality of admitted students. Some of the diversity in the curricular preparation of applicants may be reduced in future years by greater coordination between high school and university personnel. Greater standardization may also be achieved in school evaluation of achievement and ability, by the same process. There is, however, little likelihood of a complete

reversal of the policies of the last ten years, i.e., a return to common grade 13 curriculum and province-wide common examinations in each subject area. The fundamental changes which have occurred, a wider range of pre-university curricula and of ability levels appear to be here to stay, despite some movement back in the direction of a required core curriculum. The diversity, therefore, must be accommodated by the universities in greater efforts in the areas of student evaluation and classification.

Into this scenario of wider classification and evaluation activities by universities, the current variability in the marking standards employed by high schools enters as a compounding problem, particularly in some of the more sensitive classifications, e.g. in the selection of students for places in oversubscribed institutions or programmes. Nonetheless, judging from the findings of several studies to date[10] which have analysed the predictive validities of high school marks in Ontario, these marks maintain a high correlation with subsequent university evaluations of academic performance. The correlations vary, of course, with the measures correlated, especially with degree of reliability obtainable with repeated measures of similar abilities, and with the homogeneity of the groups upon which the calculations have been made (same programmes, sex, etc.). Overall, however, they continue to outperform (as predictors) alternative measures, such as standardized test scores.

---

[10] For a discussion of these sources and comparisons among their findings consult: R. Traub, et al, Secondary-Postsecondary Interface Project II, Nature of students, Vol. 1, 1976, pp. 160-161

The universities, therefore, are put in something of a quandary. Should the superior predictive validity of high school marks be sacrificed for the sake of the unbiased but less valid achievement test scores? Both measures are "unfair" to some students, but for different reasons. Hence the current clamour for "correcting" of an applicant's year 5 marks according to the past performance of his/her high school's registrants in the programme or university concerned. Such corrective adjustments are by no means a straightforward matter. Considerable care has to be exercised to avoid making statistically unwarranted corrections on the basis of small unrepresentative school samples.

Further, it would be most unfortunate if Ontario universities were to plunge headlong into achievement testing programmes and find, as with SACU experience of the early '70s, that they are unsure of what to do with the scores obtained. For this reason, subsequent analyses will not emphasize the statistical characteristics of the student populations sampled in the COU experimental projects, but will attempt to identify what can be done with the measures available.

The primary population to which analysis is directed is the cohort of current year 5 graduates. The reasons for this design are:

- it avoids the possible confounding effects of mark inflation from prior year Ontario high school entrants;

- this is the group directly concerned in the standardization of high school marks;

- this group is closely comparable to one of the groups tested in the Interface study;

- identification of membership in this group, on the basis of data elements available in the study, seems to be more reliable than for students coming from other sources.

4.2 Alternate Academic Performance Measures

4.2.1 Interim-Final Year 5 Marks

Since Ontario universities rely heavily upon year 5 interim marks in assessing an applicant's suitability for admission to a university or programme, it is important to them that these are reliable and unbiased estimates of final marks obtained.

Table 11 shows interim/final regressions at three levels of aggregation, overall average, major subject group and subject level. These values are based upon all current year 5 graduates registering in first year programmes in fall 1975 at Queen's, McMaster and Waterloo.

Table 11

Relationships between Interim* and Final Year 5 Averages
for Direct Entrants in Fall '75 to McMaster, Queen's and Waterloo

Criterion	Predictor Variable	Regression Coefficient	S.E. of Regression Coefficient	Correlation Coefficient	Standard Error of Estimate	No. of Cases
Yr 5 Final Average Best 6 of 6 Credits	Yr 5 Interim* Average	0.940	0.004	0.956	2.52	6617
	- Constant	4.306				
Yr 5 Final Average Best 6 of 6 Credits	Yr 4 Average	0.857	0.011	0.800	5.14	3749
	- Constant	9.785				
Yr 5 Final Maths Average	Yr 5 Interim* Maths Average	0.935	0.004	0.943	4.01	5974
	- Constant	5.076				
Yr 5 Final Science Average	Yr 5 Interim Science Average	0.927	0.005	0.933	3.68	6098
	- Constant	5.222				
Yr 5 Final Humanities** Average	Yr 5 Interim Humanities Average	0.915	0.005	0.926	3.37	6052
	- Constant	6.415				
Yr 5 Final English Average	Yr 5 Interim English Average	0.900	0.006	0.909	3.82	5585
	- Constant	7.416				
Yr 5 Final History Average	Yr 5 Interim History Average	0.898	0.008	0.920	3.83	2121
	- Constant	7.826				
Yr 5 Final Functions and Relations	Yr 5 Interim Functions & Relations	0.902	0.005	0.926	4.70	5081
	- Constant	7.802				

* Includes Preliminary and Early Final Marks
** English, Foreign Languages, Creative Arts, History

Interim marks include EARLY FINAL marks obtained from semestered high schools as well as preliminary final marks. The results in Table 11 show that year 5 final average of the best six credits, is predictable from the corresponding interim value to within a standard error of 2.5 marks (on scale of 0 to 100) i.e. given the usual assumptions concerning the distribution of errors in predictions, the actual final average will fall within plus or minus 2.5 marks of the predicted final mark for 68% of applicants. A high degree of association also holds down to the single course level of aggregation. The interim year 5 averages are poorer predictors of first year university performance than are final averages; the attenuation in predictive validity coefficients being approximately 5%.

4.2.2 Comparison of First Year University Credit Weighted Average from University Data Element with Average Computed from Course Marks

For McMaster (1975 and 1976) and Waterloo (1975), the university student record contained an end of session programme average. Initial analyses showed that a substantial number of students at McMaster, fewer at Waterloo, were not assigned such an average. Further cross-tabulation demonstrated that these students were in general recorded as either having withdrawn, failed the year, or as not having completed the year. Credit weighted sessional averages were also computed by COU from course marks, including all courses for which a numeric mark was present on a student's record. Because no students were excluded, the computed averages showed greater variability and were available for more students than did the average from university sources. In the latter case the marks distribution was essentially truncated at about the 40% level. Since this variable was to play an important part in subsequent analyses as a criterion measure it was important to know how well the two sources agreed when both values were available. Table 12 gives comparative statistics for McMaster and Waterloo.

There is clearly good, though not perfect agreement between the two sources. To avoid overestimating correlations by using a truncated criterion measure, the version computed from course marks was adopted for subsequent analyses, even though the suspicion exists that some of the lower averages are not truly representative of

Table 12

Relationships between University Sessional Credit Weighted Averages and those Computed from Course Marks for Direct Year 5 Entrants in 1975

Criterion	Predictor	Regression Coefficients	S.E. of Coefficient	Correlation Coefficient	Standard Error of Estimate	No. of Cases
Final 1st year Sessional Average Computed from Course Marks	Final 1st year Sessional Ave. Supplied by University*	1.113	0.002	0.990	1.64	4454
	- Constant	-9.352				

*Not available for Queen's Student Records

student abilities but arise merely from incomplete records.

Analyses relating to student classification decisions follow.

4.3 Validation of Achievement Test and Year 5 Measures for Student Classification Decisions

Although no attempt will be made to delve deeply into current thinking in the area of student classification and assessment, some concepts and general principles can be briefly noted in order to provide clues as to appropriate ways of validating tests for particular classification decisions.

Test scores and other academic performance measures can be broadly categorized as assessing;

a) Current Achievement - knowledge of subject matter or mastery of skills,

b) Potential Achievement - measures of aptitudes or general characteristics which affect learning speed, difficulty level, etc.

Information obtained concerning current achievement is used in making decisions (by or on behalf of students) about the appropriate entry point for that student into an instructional sequence, in order to avoid gaps

in, or overlapping of, instruction. Examples are:

- assignment to an appropriate level course in a sequence;
- exemption from introductory or regular course work; advanced standing;
- assignment to remedial instruction.

Information concerning potential achievement is used in making decisions about the grouping of students by their potential for responding to alternate educational treatments. Examples are:

- assignment to alternate instructional methods suited to student characteristics;
- placement in regular or accelerated programmes, matching rate of instruction to potential learning speed;
- selection of students for programmes or courses on the basis of their differential potentials for success, i.e. aptitudes.

In order to validate a test for any of these or other classification decisions, the type of assessment - current achievment or potential achievement - has to be determined and the appropriate criterion measures specified.

The following is an analysis of the more important student classification decisions and a discussion of the relevant validation procedures for measures which might be used in making the decisions.

4.3.1 Selection

The use made here of the term selection is confined to institutional and programme admission decisions. In this respect, selection differs from placement, where students are positioned at the appropriate point in an instructional sequence.

Admissibility is defined disjunctively in most programmes or institutions, a variety of "equivalent" academic qualifications being recognized. For the bulk of applicants to Ontario universities, admissibility is generally established by successful completion of the SSHGD programme and attainment of a minimum average of 60% on the best 6 credits obtained in the final year 5. No specific "core" of subjects is required at the present time (nor at the time of the COU studies), either for the SSHGD programme or for inclusion in the minimum average. At the university programme level, however, further specialized requirements, appropriate to the programme are often added to the general requirements to determine admissibility for Ontario year 5 graduates.

The additional requirements may be simply particular year 5 subject prerequisites or may further specify a minimum average on some subset of year 5 courses.

When there are more admissible applicants for a programme than places available to be filled, some means of choosing among the applicants must be adopted. In theory, the selection criteria used may be random or based upon egalitarian, or meritocratic principles, or a mixture of two or more of these. We are assuming in what follows that assessment of prior academic performance is generally used for selection in Ontario and furthermore that the selection measure is validated by its efficiency in predicting subsequent success in the programme concerned.

With respect to selection at the programme level then, we are looking for assessment measures with good predictive validity using some overall measure of ultimate success in that programme as a criterion. The data acquired in the COU studies limit the choice of criterion to first year average, or result of session, e.g. proceed, fail, withdraw.

Using first year average as the criterion measure, analyses have been done on the COU data, by programme of latest registration, to determine:

- the predictive validity of year 5 final average;
- the predictive validities of the achievement test measures used in the COU experimental administrations

- the incremental predictive validity obtainable from composites of test scores and year 5 marks.

There are many ways of presenting predictive validity results, some more concise than others. In the University Records Survey (URS) of the Interface Study[11], the researchers chose to summarise predictive relationships across all universities' programmes and high schools by the use of multiple linear regression analyses using dummy variables for each university, programme and school. This approach has the virtues of conciseness and generality and makes good use of small sample sizes in individual categories. A modified version of this approach has been applied to the COU data as a means of analysing the effects of high school variations in marking standards. However, for several reasons a breakdown by programme and university is presented here.

- Much larger samples were available to the present study and consequently statistically significant results could be obtained at the programme level of disaggregation;

- Several previous Ontario studies had employed a programme (or faculty) breakdown in reporting predictive validities;

---

[11] Ibid., pp. 156-164

- the individual idiosyncracies of universities or programmes can be taken into account in interpreting particular values, e.g. degree of selection.

Another difference between the method of analysis adopted here and that of the University Records Survey concerns the definition of university programme. In the URS a student was assigned to one of four "programmes" - arts/humanities, social sciences, sciences, professions. The assignment was based upon a count of the number of courses taken in each programme category and the programme was determined by the category having the highest count.

In contrast, most COU analyses were done using institutional definitions of programme. In some instances institutionally defined programmes were amalgamated where admissions and core first year course requirements were substantailly the same for the programmes concerned. Thus the raw data for assessing predictive validities are presented by programme and university in Tables 13.1 to 13.3 for the 1975 participating institutions and in Tables 14.1 and 14.2 for those of 1976 for which university data were available for analysis.

A variety of predictors are shown including composites of achievement test subtest scores[12], total scores and year 5 average. No unique total

---

[12] As used here and in the Tables, the term composite connotes the weighted sum of predictor scores which best predicts the criterion (the programme average in this case). The correlation coefficient shown for a composite predictor is therefore a multiple correlation coefficient.

Table 13.1

Means, Standard Deviations & Correlations for Year 5 Average (1975-76) and University First Year Programme Average

Programme	Correlation Coefficient Yr 5 Ave. with 1st Yr Ave.	Standard Error of Estimate	First Yr Univ. Ave. Mean	First Yr Univ. Ave. S.D.	Year 5 Average Mean	Year 5 Average S.D.	No. in Sample *
McMaster							
Humanities I	0.62	8.8	67.7	11.3	75.1	7.9	225
Social Science I	0.54	9.1	62.5	10.8	69.5	6.8	479
Natural Science I	0.68	11.5	62.3	15.7	77.0	8.9	487
Engineering I	0.67	10.7	64.4	14.4	77.6	8.4	300
Business I	0.49	10.1	63.8	11.6	70.3	7.6	325
Physical Educ. I	0.58	8.8	60.8	10.7	69.0	7.7	244
Nursing	0.57	5.8	71.0	7.0	83.4	3.7	54
Queen's							
Arts	0.54	6.3	67.9	7.5	76.9	5.6	534
Science (Hons.)	0.66	8.4	68.9	11.3	80.9	6.4	393
Applied Science	0.66	7.8	60.9	10.4	78.6	7.7	320
Commerce	0.68	5.7	69.5	7.8	77.9	5.5	122
Arts-Physical Educ.	0.68	5.6	64.0	7.5	74.8	5.9	80
Waterloo							
Arts (General)	0.58	7.1	69.4	8.7	69.3	7.9	492
Science	0.70	8.9	66.5	12.5	72.7	9.3	414
Mathematics	0.65	8.8	69.5	11.5	80.2	8.0	679
Engineering (General & Chemical)	0.62	8.1	65.6	10.3	77.4	6.7	560
Systems Design	0.78	6.7	72.0	10.7	78.7	8.4	46
Human Kinetics & Leisure Studies	0.64	6.0	70.2	7.8	71.3	6.7	264
Geography	0.64	5.6	70.6	7.3	70.7	7.1	144
Urban Planning	0.70	3.8	77.4	5.3	78.2	6.8	53
Architecture	0.65	4.6	73.4	6.0	79.8	7.2	39
Man Environment	0.41	9.8	74.4	10.7	71.8	6.0	59

* All registrants coming directly from Ontario Year 5 are included

Table 13.2

Means, Standard Deviations & Correlations for CTEL, Year 5 Average (1975-76) and University First Year Average

Programme	Correlations of Final 1st Year Programme Average With: CTEL Composite	CTEL Total	Year 5 Ave.	Year 5 Ave. and CTEL Comp.	CTEL Total Mean	CTEL Total S.D.	Year 5 Ave. Mean	Year 5 Ave. S.D.	No. in* Sample
McMaster									
Humanities I	0.24	0.20	0.61	0.62	44.9	13.2	76.1	7.8	162
Social Science I	0.32	0.28	0.56	0.57	37.4	10.6	70.1	6.9	269
Natural Science I	0.29	0.28	0.64	0.65	42.8	13.1	77.8	8.7	358
Engineering I	0.21	0.16	0.67	0.68	37.4	12.6	78.1	8.6	231
Business I	0.37	0.37	0.58	0.60	36.9	10.9	71.1	7.9	188
Physical Educ. I	0.41	0.41	0.57	0.60	34.2	10.2	68.9	7.7	127
Nursing	0.64	0.61	0.51	0.67	43.5	11.2	83.9	3.6	45
Queen's									
Arts	0.40	0.39	0.57	0.59	50.2	11.8	77.7	5.8	116
Waterloo									
Arts (General)	0.37	0.37	0.59	0.61	40.6	11.9	70.5	8.1	226
Science	0.34	0.33	0.72	0.73	40.3	12.8	74.1	8.9	213
Mathematics	0.37	0.37	0.65	0.66	42.3	13.6	81.1	7.9	448
Engineering	0.45	0.44	0.66	0.70	40.9	12.0	78.6	7.0	300
Systems Des.	-	0.37	0.82	-	42.8	14.2	80.7	7.9	25
Human Kinetics & Leisure Studies	0.37	0.36	0.70	0.71	38.0	10.6	73.0	6.8	103
Geography	0.45	0.44	0.68	0.71	42.1	10.7	73.1	7.6	66
Urban Planning	0.50	0.39	0.56	0.69	43.2	12.8	79.1	6.5	33
Architecture	-	0.35	0.64	0.67	48.1	13.0	80.6	7.1	30
Man Environment	0.57	0.19	0.49	0.67	36.4	9.4	73.0	5.4	26

*Only CTEL Volunteers coming directly from Ontario Year 5 are included.

Table 13.3

Simple and Multiple Correlation Coefficients for CMPT and Year 5 Average with First Year University Final Programme Average (1975-76)

Programme	CMPT Vol. Group	Correlations of First Year Average with: CMPT section or Composite	CTEL TOTAL	CMPT & CTEL TOTAL	Year 5 Average	Yr. 5 Ave. & CMPT	Yr. 5 Ave. & CMPT & CTEL	Means & Standard Deviations of Predictors: CMPT Mean	CMPT S.D.	CTEL TOT. Mean	CTEL TOT. S.D.	YEAR 5 Ave. Mean	YEAR 5 Ave. S.D.	No. in Group
McMaster: Natural Science	A1	0.49	0.23	0.50	0.63	0.67	0.68	15.1	5.0	42.5	13.2	78.0	9.1	133
	A2	0.43	0.35	0.48	0.66	0.68	0.68	9.7	3.7	43.1	12.9	77.8	8.6	206
	B1	0.44	0.29	0.47	0.64	0.67	0.67	12.9	4.8	43.0	13.0	77.8	8.8	310
	B2	0.36	0.32	0.42	0.65	0.67	0.67	13.8	4.2	43.8	13.0	78.5	8.8	223
	B3	0.38	0.28	0.41	0.64	0.65	0.65	6.3	3.5	40.6	12.7	76.8	8.6	142
	A1 B1 B2	0.45			0.67	0.72		44.4	12.6			79.4	9.2	73
	A2 B1 B2	0.46			0.62	0.65		35.8	10.5			78.1	8.6	132
Engineering	A1	0.56	0.16	0.58	0.62	0.66	0.66	15.7	4.2	38.5	12.8	79.2	8.3	113
	A2	0.47	0.20	0.47	0.76	0.77	0.77	10.1	3.8	36.9	11.9	77.0	8.7	111
	B1	0.55	0.17	0.56	0.70	0.73	0.73	14.3	5.0	37.4	12.5	77.9	8.8	199
	B2	0.43	0.21	0.45	0.70	0.72	0.72	15.9	4.0	37.5	13.0	79.0	8.5	158
	B3	0.48	0.39	0.40	0.54	0.55	0.55	7.6	3.5	39.1	10.5	77.1	8.3	88
	A1 B1 B2	0.67			0.75	0.79		48.0	11.9			79.3	8.2	75
	A2 B1 B2	0.56			0.78	0.80		39.4	11.8			77.6	9.1	69
Business	A1	0.52	0.40	0.54	0.59	0.64	0.64	11.3	4.2	37.3	11.9	71.5	8.5	81
	A2	0.29	0.36	0.40	0.58	0.61	0.62	7.7	2.8	36.8	10.3	71.4	7.7	92
	B1	0.33	0.37	0.41	0.63	0.63	0.63	10.1	4.7	37.4	11.0	71.6	8.2	155
	B2	0.28	0.39	0.40	0.57	0.59	0.60	10.8	3.9	36.4	10.6	71.3	7.9	115
	B3	0.36	0.36	0.45	0.51	0.53	0.54	5.5	3.1	37.3	11.9	71.3	8.1	76
	A1 B1 B2	0.52			0.74	0.74		29.9	12.6			71.6	8.4	43
	A2 B1 B2	0.42			0.62	0.67		27.6	9.0			71.0	7.6	61
Queen's: Applied Science	A1	0.44	-	-	0.63	0.64	-	18.0	4.1	-	-	81.0	7.7	41
	A2	0.69	-	-	0.65	0.76	-	15.1	4.2	-	-	80.4	8.0	76
	B1	0.54	-	-	0.65	0.68	-	19.2	4.3	-	-	80.7	8.0	106
	B2	0.68	-	-	0.64	0.76	-	18.1	3.8	-	-	80.1	7.6	65
	B3	0.56	-	-	0.59	0.68	-	11.2	4.2	-	-	80.7	7.8	65
	A1 B1 B2	0.79	-	-	0.58	0.83	-	54.8	11.0	-	-	80.4	8.0	19
	A2 B1 B2	0.74	-	-	0.72	0.82	-	52.3	11.4	-	-	80.0	8.0	35

Table 13.3 cont.

Simple and Multiple Correlation Coefficients for CMPT and Year 5 Average with First Year University Final Programme Average (1975-76)

Programme		Correlations of First Year Average with:						Means & Standard Deviations of Predictors						
		CMPT section or Composite	CTEL TOTAL	CMPT & CTEL TOTAL	Year 5 Average	Yr. 5 Ave. & CMPT	Yr. 5 Ave. & CMPT & CTEL	CMPT Mean	CMPT S.D.	CTEL TOT. Mean	CTEL TOT. S.D.	YEAR 5 Ave. Mean	YEAR 5 Ave. S.D.	No. in Group
Queen's Commerce cont.	A1	0.34	-	-	0.57	0.57	-	16.0	5.0	-	-	78.4	5.2	26
	A2	0.51	-	-	0.63	0.68	-	10.0	3.3	-	-	77.4	5.2	60
	B1	0.42	-	-	0.63	0.64	-	13.6	4.2	-	-	77.6	5.4	79
	B2	0.44	-	-	0.44	0.51	-	14.4	3.9	-	-	77.0	5.1	54
	B3	0.35	-	-	0.79	0.79	-	8.4	3.9	-	-	79.0	5.2	35
A1 B1 B2		-	-	-	-	-	-	-	-	-	-	-	-	-
A2 B1 B2		0.52	-	-	0.46	0.60	-	36.2	8.4	-	-	76.7	5.0	39
Waterloo: Science	A1	0.55	0.39	0.59	0.71	0.72	0.72	16.2	4.9	42.7	14.6	73.9	9.5	71
	A2	0.26	0.31	0.36	0.73	0.73	0.74	11.4	3.5	39.6	11.9	74.5	8.6	134
	B1	0.42	0.37	0.49	0.74	0.74	0.75	14.5	4.7	40.1	13.0	74.6	9.0	191
	B2	0.32	0.29	0.36	0.68	0.68	0.69	14.4	4.3	40.2	12.7	74.7	8.6	162
	B3	0.43	0.35	0.46	0.73	0.75	0.75	8.1	4.0	41.6	13.3	72.1	9.7	57
A1 B1 B2		0.61			0.72	0.73		46.9	13.0			74.3	9.1	49
A2 B1 B2		0.26			0.68	0.71		39.4	11.8			75.4	8.3	105
Mathematics	A1	0.43	0.32	0.47	0.59	0.61	0.62	18.9	3.7	43.2	14.1	80.8	7.7	236
	A2	0.44	0.43	0.51	0.69	0.70	0.71	13.6	4.1	41.3	12.9	81.7	8.0	204
	B1	0.50	0.37	0.53	0.64	0.66	0.67	18.1	4.5	42.7	13.6	81.5	7.8	406
	B2	0.47	0.38	0.51	0.64	0.67	0.68	17.9	4.1	42.0	13.6	81.5	8.0	287
	B3	0.36	0.32	0.42	0.62	0.64	0.64	9.0	3.6	42.1	13.6	79.9	7.6	187
A1 B1 B2		0.52			0.60	0.63		55.7	11.1			81.6	8.0	148
A2 B1 B2		0.61			0.69	0.74		51.6	7.6			83.0	11.7	108
Engineering	A1	0.56	0.46	0.63	0.74	0.78	0.79	17.3	4.2	40.9	12.6	78.3	7.0	152
	A2	0.48	0.40	0.53	0.58	0.64	0.65	12.8	3.5	40.7	11.3	78.6	7.0	149
	B1	0.53	0.43	0.58	0.67	0.70	0.71	17.0	4.1	40.8	12.1	78.5	6.9	274
	B2	0.44	0.44	0.52	0.68	0.71	0.73	17.5	3.4	41.0	11.4	78.9	7.1	210
	B3	0.43	0.47	0.53	0.61	0.63	0.67	8.8	3.5	40.2	12.8	78.1	6.8	113
A1 B1 B2		0.65			0.77	0.82		52.0	10.2			78.6	7.3	108
A2 B1 B2		0.58			0.60	0.67		47.7	8.7			78.9	6.9	84

Table 14.1

Means, Standard Deviations and Correlations for RCLAT, Year 5 Average (1976-77) and University First Year Programme Average

Programme	Correlations of Final 1st Year Programme Average With: RCLAT Composite	RCLAT Total	Year 5 Ave.	Year 5 & RCLAT	RCLAT Total Mean	RCLAT Total S.D.	Year 5 Average Mean	Year 5 Average S.D.	No in Sample*
Carleton									
Arts (Pass & Hons)	0.32	0.31	0.62	0.62	17.5	6.9	69.8	7.8	754
Science (Pass & Hons)	0.46	0.46	0.71	0.73	19.0	7.3	74.6	9.7	195
B. Engineering	0.27	0.21	0.59	0.61	16.8	7.3	74.4	8.5	148
Industrial Des.	-	-	0.46	0.46	16.4	9.3	73.3	6.2	30
B. Commerce	0.34	0.31	0.62	0.64	17.1	6.6	73.2	6.6	175
B. Journalism	0.28	0.17	0.55	0.60	23.0	5.7	77.7	6.5	111
B. Architecture	0.58	0.57	0.47	0.65	20.9	7.7	85.5	5.1	42
McMaster									
Humanities I	0.25	0.24	0.57	0.57	19.1	7.1	73.7	7.7	228
Social Science I	0.34	0.33	0.57	0.59	15.6	7.3	68.8	6.6	468
Natural Science I	0.39	0.38	0.69	0.70	18.3	7.0	78.4	8.3	493
Engineering I	0.28	0.26	0.76	0.77	17.3	7.6	78.7	8.1	266
Business I	0.24	0.24	0.58	0.58	15.2	6.8	71.9	7.7	323
Physical Ed. I	0.34	0.28	0.52	0.56	14.3	6.1	69.2	6.3	182
Nursing	0.34	0.26	0.33	0.41	21.6	5.4	83.6	3.8	51

*Only RCLAT writers coming directly from Year 5 are included.

Table 14.2

Means, Standard Deviations and Correlations for RCLAT, Essay, Year 5 Average (1976-77) and University First Year Programme Average

Programme	Correlations of Final First Year Average with: RCLAT Total	Essay	RCLAT & Essay	Year 5 Average	Year 5 & Essay	Year 5 & Essay & RCLAT	RCLAT Total Mean	RCLAT Total S.D.	Essay Mean	Essay S.D.	Year 5 Average Mean	Year 5 Average S.D.	No. in Sample*
McMaster													
Humanities I	0.43	0.34	0.46	0.66	0.66	0.67	20.1	7.5	7.1	1.3	74.2	8.2	84
Social Science I	0.47	0.34	0.52	0.56	0.59	0.65	15.5	7.3	6.1	1.3	69.2	7.0	144
Natural Science I	0.41	0.35	0.46	0.72	0.74	0.75	18.3	7.2	6.4	1.6	78.3	8.0	144
Engineering I	0.32	0.29	0.43	0.76	0.77	0.78	17.2	7.9	6.0	1.5	78.4	8.4	82
Business I	0.09	0.22	0.27	0.54	0.55	0.57	15.4	7.1	6.0	1.4	72.0	7.1	82
Physical Educ.I	0.24	0.22	0.43	0.65	0.66	0.70	14.9	5.9	6.3	1.1	70.6	6.4	48
Nursing	0.26	0.28	0.77	0.68	0.69	0.85	21.6	5.4	6.9	1.0	83.6	3.8	17

* Only students who wrote an Essay and the RCLAT, and come directly from year 5 are included.

score was available for CMPT since students were allowed to select the three sections of their choice from the five sections in the CMPT battery.

On the evidence presented in Tables 13.1 to 14.2 the following general observations may be noted:

a) High school average remains the single best predictor of overall first year university success, and at levels comparable to previous findings of the late '60s and early '70s. The degree of prediction obtained varies from university to university, and from programme to programme. In general year 5 average has higher predictive validity for science and engineering programmes than for arts and other programmes. There are a number of factors which contribute to this commonplace finding in Ontario and other jurisdictions.

(i) The science, mathematics and engineering registrants are more homogeneous in the subject composition of their year 5 average, i.e. most have at least 2 mathematics and 2 science credits as a "core". This is largely attributable to the more specialized subject requirements for admission to these programmes.

(ii) most of the subjects in the average are repeated measures of quantitative skill, and the average is thus more reliable than an average based upon diverse abilities.

(iii) the criterion measure, e.g. grade point average in engineering is similarly reliable (repeated measures of quantitative ability), and similarly homogeneous with regard to the composition of the average, i.e. engineering students take a common set of courses in first year with little elective discretion being allowed.

(iv) mathematics and science marks spread students more widely over the score range, and therefore discriminate better between students (witness standard deviations in year 5 and university subject averages).

b) The addition of test information to high school average has only a small effect upon predictive validity for university first year programme average. The effect is more pronounced for the mathematics test (CMPT) than for either English test (RCLAT or CTEL), but small nevertheless.

Some apparent anomalies occur where the predictive validity of the year 5 average is underestimated owing to the severe restriction in the range of this measure caused by its use in selecting for admission. This underestimation follows from the explicit use of the year 5 average in the admission process and a consequent reduction of the variability of year 5 marks of admitted students as compared to the variation which would be apparent in the total applicant pool. In Table 13.2, for example, the standard deviation of the year 5 average for students admitted to the McMaster nursing programme is only 3.6 compared to values in the region of 8.0 for most other programmes. The CTEL total score is much less affected by the selection process, the standard deviation being reduced from about 13.0 (for groups with similar means) to 11.2 for nursing students. Some other programmes showing indications of underestimation of year 5 average predictive validity are McMaster social science, Queen's arts, Waterloo engineering, urban planning and man environment. Another point to be considered when evaluating the predictive validities of these tables is that no great reliance should be placed upon estimated validity coefficients which are based on small numbers of students. Such estimates can fluctuate widely from sample to sample.

c) Achievement test scores taken alone, while not possessing the degree of predictive validity of year 5 averages, are significantly positively correlated with final first year university programme

average. In Table 13.3, it may be seen that CMPT subtest scores are generally more highly correlated with first year averages in engineering, science, and mathematics programmes than is the CTEL total score. Subtests A1-Algebra and B1-Trigonometry and Functions are, by and large, the best predictors among CMPT subtests, except in the case of the Queen's applied science programme where A2-Calculus and B2-Synthetic and Coordinate Geometry are most effective (only 41 students elected to take the A1-Algebra subtest in this programme however). The composite predictors formed from three CMPT subtests raise predictive validities substantially for some programmes, but not so far as to challenge the superiority of the year 5 average in the programme concerned. Predictive validities of CTEL and RCLAT range from 0.24 to 0.46 (ignoring the smaller samples). The highest values are reached in Waterloo's engineering (0.45) and geography (0.45) and McMaster's physical education (0.41) programmes for CTEL; and in Carleton's science (0.46), architecture (0.58) and McMaster's science (0.39) programmes for RCLAT. The combination of RCLAT total score and the essay score raised the predictive validity (over either score alone) for programme averages in humanities, social science, natural science, and engineering at McMaster (see Table 14.2). The composite gave predictive validities between 0.43 and 0.52 for these programmes.

d) Composites of CMPT subtest score and CTEL scores show modest improvements in prediction over the CMPT score alone in all programmes. The improvement is most marked for Waterloo's science, mathematics and engineering programmes where the average increase in validity coefficient is approximately 6%.

What, then, do these figures say about the relative and differential utilities of standardized tests and high school averages for admissions decisions, in particular for selection to limited enrolment programmes where differences in high school marking standards make the use of unadjusted high school averages unfair to some students? How much in selection efficiency is lost by adopting unbiased measures, such as the achievement tests administered in the COU studies?

The main options for using tests and high school averages for selection are as follows:

i) Use a standardized test or a battery of such tests, ignoring high school performance data altogether. This would be perfectly fair with respect to variations in school marking standards but predictive validities are substantially lower than can be obtained <u>with</u> high school measures or composites of test scores and high school measures. This option has the advantage, however, of allowing for direct comparisons between Ontario year 5 graduates and other applicants.

(ii) Use a selection measure which combines the use of unadjusted high school and standardized test measures. The hope here would be for the new measure to be sufficiently unbiased with respect to variations in high school marking standards that it could be considered "fair". Unfortunately, as the analysis of the next section seems to indicate, a single achievement test score will not accomplish a significant reduction in the advantage conferred by attending a school with lenient marking standards.

(iii) Use a two-stage selection strategy in which high school data are used for the initial screening of applicants whose marks are high enough for admission regardless of variations in standards, and then give standardized tests to the group whose marks could be affected by such variation. This method still penalizes applicants from the schools with more severe marking policies by making them more likely to have to take a test, where "all is on the hazard".

(iv) Use achievement test scores from previous years (most appropriately given to new registrants at some time soon after registration) in order to statistically adjust for differences in high school marking scales. There is a serious question here as to whether the test scores are valid surrogate measures of high school achievement, particularly when only a small sampling of subjects is included in the test battery. Also, correcting in this manner is hardly warranted before schools have had an opportunity to

familiarize themselves with the content of the tests and adjust their curricula accordingly. Otherwise, differences in performance on the tests, between the schools, are likely to reflect differences in curricular emphasis as well as in marking standards.

(v) Ignore achievement tests altogether and rely upon high school measures only. A variation in this method would be to adjust high school marks using historical university academic performance data on each school's graduates.

(vi) Combine adjusted high school marks with achievement test scores to obtain a maximum improvement in selection validity.

In the light of the results presented in Tables 13.1 through 13.3, Option i, selecting students on the basis of one or two achievement test scores alone, does not merit serious consideration. A considerable loss in SELECTION EFFICIENCY would be incurred in most programmes if high school data were ignored. The effect of such a reduction in predictive validity would vary according to the

SELECTION RATIO of the programme concerned.[13] Where almost all applicants are admitted, the measure used to select the "best" students (as inferred from their subsequent performance in the programme) makes little difference. In the case of limited enrolment programmes, however, where low selection ratios hold, substantial changes in selection efficiency can occur if there are differences between the predictive validities of the measures used for selection (see Table A5).

Incidentally, Table A5 shows why decreasing the selection ratio in a programme is more wasteful of the potential talent in the applicant pool than increasing the predictive validity of the selection measure used. A reduction in the selection ratio invariably leads to a reduction in the proportion of the "best" students selected, whereas increasing the predictive validity of the selection measure invariably leads to a higher expected proportion of the "best" students being selected. Admittedly, by decreasing the selection ratio the proportion of selected students exceeding a fixed criterion

[13] The term selection ratio is defined as the ratio of the number of students selected to the number eligible for selection. In the context of selection to limited enrolment programmes, the selection ratio becomes the ratio of the number of places to be filled to the total number of eligible applicants.

The term selection efficiency is defined here as the proportion of the students selected at a given selection ratio who would also have been selected by the criterion measure had it been available. In other words the selection efficiency indexes the efficiency with which a predictor selects the "best" students from among those eligible for selection.

An expanded description of these two terms can be found in Appendix A5.

score would increase, as would the average criterion score obtained by this smaller group. However, a larger proportion of the best students at the selection ratio used would have been rejected. It is possible that the composite predictors formed from several standardized test scores might show predictive validities which are comparable to Ontario year 5 averages. However, we cannot on the basis of the COU test administrations, comment on more than two tests given to any group of students. Furthermore, none of the achievement tests administered were designed specifically to be predictors of programme success.

The increments in predictive validities obtained for composites of year 5 average and achievement test scores, are not sufficient to raise programme selection efficiencies very much above what is provided by the year 5 average alone. Table A5 reveals that at a selection ratio of 50%, an increase in the predictive validity of the selection measure from 0.65 to 0.70 results in an improvement in selection efficiency from 73% to 75%; at a very low selection ratio, 10% for example, an improvement in selection efficiency from 43% to 47% would ensue. A selection ratio as low as 10% is unlikely to be encountered in practice however, and the incremental predictive validity of 0.05 is about as high as can be obtained from combining the available test scores with the year 5 average. It seems, therefore, that under realistic conditions option (ii) would only marginally enhance the ability of admissions officers to select the "best" students from the applicant pool.

No attempt has been made to estimate the effect on programme selection efficiency of the remaining options from the COU sampling design. For options involving corrections to year 5 marks, samples of students from each of a representative subset, but not necessarily from all Ontario high schools are required. The COU data can provide a large number of registrants, but spread thinly over most schools in the province.

The discussion of selection up to this point has been restricted to limited enrolment programmes, and consequently the only criterion measure treated has been the overall first year programme average. There are, however, other selection decisions for which this criterion is too heterogeneous to be of use. A pure criterion would be required to assess the validity of the available English measures for purposes of selecting the students most likely to succeed (or fail) in university subjects where English language skills are a major component in student evaluation at the first year undergraduate level. Since a good deal of essay writing is normally required of students taking courses in English and in history, and moreover since these two subjects are fairly heavily subscribed, they may provide some approximation to criteria of university evaluated writing ability.

Some of the drawbacks to using the university English or history average as a criterion of the writing competence of entering freshmen are as follows:

a) Students rarely take more than a single credit in English or in history in their first year, so that the average is based upon a single mark, and is in consequence a less reliable measure than a programme average (or the mathematics average criterion based on two credits).

b) Students do not all take the same course, in English (or history) and marking standards may vary somewhat even within a single institution (the exception here being the students who wrote CTEL at Queen's, all of whom were taking English 110).

c) First year English courses frequently have instruction in essay writing built into the course curriculum. This might tend to obscure differences in writing ability that may have existed between students at the start of the year. However if the tests are seen as predicting "survival" in these courses rather than level of preparation upon entry, then the objection raised above is irrelevant.

d) The students taking English and history courses are not representative of the freshmen population as a whole in terms of their English language abilities as evaluated in their year 5 English average and CTEL total score. We are dealing here with a group of students who are both self-selected and institutionally selected on their prior achievement in English language skills. In particular, it is unlikely that the students with the weakest English abilities would elect to take courses in English and history.

Tables 15.1, 15.2 and 15.3 show predictive validities of the English tests, and year 5 English average together with incremental predictive validities of the composite measure formed by combining the year 5 English average with achievement sub-test scores, for three criterion measures; namely university English, history and mathematics average. The latter criterion is included to give some perspective to the degree of prediction estimated for English and history grades by the various predictors presented.

In all but one instance, the year 5 English average was the best single predictor of the first year university English or history mark. The predictive validities of CTEL and RCLAT were similar to each other though low in comparison with year 5 English. Unfortunately the number of freshmen who wrote the Writing Test and also took a course in English or in history, was rather small (see Table 15.3). The results suggest that a composite of RCLAT and Essay scores was equally good in predicting first year university English and history grades, as a year 5 English grade.

Modest improvements in predictive validities (over year 5 English alone) were obtained from composites of year 5 English average and the English Achievement test scores. Most increments in predictive validity fell in the range .02 to .05 on base validities of from 0.39 to 0.57. Incremental predictive validities displayed in Table 15.3 for RCLAT and Essay scores show the latter to be marginally better predictors

Table 15.1

Means, Standard Deviations & Correlations for CTEL, Year 5 English (1975-76) and University First Year Subject Averages

First Year Subject	Correlations of 1st Year Subject Average With: CTEL Composite	CTEL Total	Year 5 English	Year 5 English & CTEL Comp.	CTEL Total Mean	CTEL Total S.D.	Year 5 English Ave. Mean	Year 5 English Ave. S.D.	No in Sample
McMaster									
English	.37	.33	.47	.50	42.2	12.2	73.4	8.2	330
History	.42	.40	.52	.55	41.3	11.5	72.8	8.1	223
Mathematics	.18	.13	.30	.31	40.8	12.5	71.6	9.5	667
Queen's									
English	.42	.41	.53	.58	50.1	12.1	80.0	6.2	116
History	.47	.40	.29	.49	50.9	11.8	80.4	5.7	79
Mathematics	-	-	-	-	-	-	-	-	12
Waterloo									
English	.38	.37	.41	.46	42.1	12.0	71.9	8.8	296
History	.20	.13	.43	.45	42.8	11.3	71.1	9.3	136
Mathematics	.33	.32	.30	.38	42.8	12.4	71.1	8.5	724

Table 15.2

Means, Standard Deviations and Correlations for RCLAT, Year 5 English Average (1976-77) and University First Year Subject Averages

First Year Subject	Correlations of 1st Year Subject Average With: RCLAT Composite	RCLAT Total	Year 5 English	Year 5 English & RCLAT	RCLAT Total Mean	RCLAT Total S.D.	Year 5 English Mean	Year 5 English S.D.	No. in Sample
Carleton									
English	.37	.37	.57	.59	19.8	6.8	73.2	9.6	462
History	.29	.23	.46	.48	19.7	6.5	72.7	9.6	340
Mathematics	.26	.25	.35	.38	17.4	7.1	69.2	10.0	587
McMaster									
English	.35	.33	.39	.44	18.0	6.8	72.8	8.1	447
History	.30	.28	.47	.48	17.2	6.6	70.5	8.3	365
Mathematics	.18	.17	.27	.28	17.5	7.1	70.9	9.7	1,021

for both the English and the history criterion. The full English battery consisting of year 5 English average, RCLAT subtest scores and Essay score had predictive validities of 0.52 for the English criterion and 0.56 for the history criterion, an overall increase in predictive validity over year 5 English of approximately 0.1. The corresponding increase in selection efficiency would be about 5% whatever the selection ratio used. The increase in selection efficiency brought about by including a single standardized score, CTEL, RCLAT or Essay, in a composite selection measure with year 5 English would amount to 3%.

The variety of first year university mathematics courses, aimed at the differing requirements of students from programmes such as business, science, engineering and the social sciences, makes an average based upon all first year mathematics grades far too heterogenous to be a useful criterion for test validation purposes. For this reason the predictive validities presented in Table 16 are based upon the results of only those students who attempted a substantial mathematics course-load (two or more full credits for McMaster and Waterloo students, and one or more for Queen's students). This restriction virtually excludes students registered in arts, business or social sciences programmes from consideration since they will rarely attempt more than one full credit in mathematics in their first year. The students included therefore are by and large registered in natural sciences, engineering or mathematics programmes where the demands made on mathematical

Table 15.3

Means, Standard Deviations and Correlations for RCLAT, Essay Scores,

Year 5 English Average (1976-77) and University First Year Subject Averages

First Year Subject	Correlations of 1st Year Subject Average with: RCLAT Composite	RCLAT Total	Essay	RCLAT & Essay	Year 5 English	Year 5 English & RCLAT	Year 5 English & Essay	Year 5 English RCLAT & Essay	RCLAT Total Mean	RCLAT Total S.D	Essay Mean	Essay S.D	Year 5 English Mean	Year 5 English S.D	No. in Sample
McMaster															
English	.40	.38	.39	.46	.41	.48	.49	.52	19.0	7.5	6.7	1.2	73.5	7.6	146
History	.41	.39	.41	.48	.51	.54	.55	.56	18.2	6.9	6.6	1.2	70.6	8.7	105
Mathematics	.21	.20	.15	.21	.27	.29	.27	.29	18.2	7.2	6.4	1.5	70.4	10.0	364

ability are greatest and where the best prepared students (in mathematics) are to be found.

As Table 16 shows, two CMPT subtests, A1-Algebra and B1-Trigonometry and Functions, were moderately good predictors of the "specialized" mathematics criterion described above. In fact their predictive validities of about 0.55 were comparable to that of a single interim year 5 mathematics course mark. The other three subtests were less effective in prediction (except at Queen's), a fact which has little bearing however on their usefulness for placement purposes. The best composite of CMPT subtest scores raised the predictive validity to the 0.60 level which is close to the value obtained for the interim mathematics average. However, the most impressive feature of the CMPT-A1 and B1 subtest scores was their ability, when combined with the year 5 mathematics average, to raise predicitive validities by approximately .05 from such high initial levels (0.61 to .68). This observation, together with the relatively low intercorrelations between CMPT subtest scores and individual year 5 mathematics course marks, i.e. 0.45 to 0.50, indicates that CMPT is not merely substituting for year 5 mathematics marks but rather is supplementing them. Even so, the increase to be expected in selection efficiency is not great, amounting to 2% when the selection ratio is 50% and 4% when the selection ratio is 10%.

4.3.2 Placement

The term placement is unfortunately somewhat ambiguous in that it is used to refer to some quite different classification decisions. The distinctions

Table 16

Means, Standard Deviations and Correlations for CMPT Subtest Scores, Year 5 Mathematics Averages and University First Year Mathematics Average* (1975-76)

University	CMPT Volunteer Group	Correlations of 1st Year Mathematics Average With: CMPT Section Score	Year 5 Math. Average	CMPT & Year 5 Math.	CMPT Section Score Mean	CMPT Section Score S.D.	Year 5 Math. Ave. Mean	Year 5 Math. Ave. S.D.	Math. Credits Attempted Year 5 Mean	Math. Credits Attempted Year 5 S.D.	Math. Credits Attempted University Mean	Math. Credits Attempted University S.D.	No. in Group
McMaster													
	A1	.56	.61	.68	16.1	4.2	81.1	10.7	2.8	0.5	2.1	0.1	193
	A2	.43	.71	.72	10.3	3.6	78.6	9.9	2.7	0.5	2.1	0.1	220
	B1	.51	.68	.72	14.2	4.8	79.5	10.4	2.8	0.5	2.1	0.1	372
	B2	.38	.68	.70	15.5	4.1	80.5	10.4	2.8	0.5	2.1	0.1	280
	B3	.37	.57	.59	7.4	3.4	79.1	10.2	2.9	0.4	2.1	0.1	169
Queen's													
	A1	–	–	–	–	–	–	–	–	–	–	–	–
	A2	.68	.61	.72	15.1	4.2	83.2	9.3	3.0	0.2	2.1	0.3	75
	B1	.57	.60	.67	19.1	4.3	83.4	9.5	3.0	0.2	2.1	0.3	107
	B2	.65	.67	.75	18.0	3.9	82.9	8.8	3.0	0.2	2.2	0.4	67
	B3	.65	.52	.70	11.3	4.2	83.1	9.1	3.0	0.2	2.1	0.3	64
Waterloo													
	A1	.53	.62	.67	18.5	3.9	82.5	8.7	3.0	0.2	2.6	0.6	397
	A2	.46	.63	.66	13.3	3.9	82.5	9.2	3.0	0.3	2.6	0.5	381
	B1	.56	.63	.68	17.7	4.4	82.7	9.0	3.0	0.3	2.6	0.5	719
	B2	.46	.63	.67	17.7	3.9	82.7	9.0	3.0	0.3	2.6	0.6	523
	B3	.47	.59	.64	9.0	3.7	81.7	8.7	3.0	0.3	2.6	0.5	313

* For McMaster and Waterloo, only students attempting 2 or more credits in University Mathematics are included. For Queen's, to increase the sample sizes, all students with more than 1 credit in University Mathematics are included.

made at the beginning of this section between the use of tests to assess current achievement and their uses in assessing potential performance distinguish several kinds of placement.

a) The assignment of students into courses or sections best suited to their current level of preparation in the subject matter concerned.

The validity of tests designed for this purpose rests mainly upon their:

- content validity as attested by course teachers (i.e. the content adequately covers course prerequisites);
- concurrent validity with some more comprehensive evaluation measure of achievement in the same subject or skill at the same level of difficulty.

Adequate reliability is also of concern for this type of testing.

b) A related placement decision is that of exempting students from the normal course work requirements on the basis of demonstrated attainment of the course objectives. Validity of a test used for exemption purposes should be established by administering the test to students who have taken the course at the completion of their study. In this way final course marks can be used as the concurrent criterion measure. Scores on the test should be strongly associated with final course marks. As with a) above, the content validity and reliability of the test should be confirmed.

c) A quite different form of placement is that of assigning or giving guidance to students as to the programme or instructional method which promises the most for their future achievement. Tests used for this purpose should display high discriminant validity with measures of success in alternate programmes or courses. In contrast to determining the validity for a test used in selection, i.e. by estimating its predictive validity for a single criterion measure of success, validity for this type of placement requires a set of tests each with higher predictive validities for a particular programme option. A test of general ability will not be of much assistance for this type of placement decision.

CMPT was designed specifically for use in placement decisions of the types described under a) and b) above. To quote from the CMPT Technical Report, CMPT "is intended for administration to students planning to study mathematics at a Canadian university and is viewed as possessing the capacity to assist in the placement of freshmen in a university's mathematics programme".

In the context of mathematics placement decisions of type a) above, content validity is established by determining to what extent CMPT items sample the domains of prerequisite skills in algebra, calculus, etc., required for first year mathematics courses in Ontario univer-

sities. No direct evidence for this kind of content validity has been assembled for CMPT in the COU studies, i.e. particular mathematics course requirements have not been analysed by the university faculty concerned and related to CMPT items. However, the National Committee responsible for CMPT content specification included a representative of Ontario universities mathematics departments.

There is a good deal more evidence for CMPT content validity[14] with reference to the Ontario secondary schools senior division mathematics programmes, i.e. years 3, 4 and 5. With the exception of section A2-Calculus, the items in each CMPT subtest are drawn from the core requirements of more than one year 5 mathematics course and also to a substantial amount of year 3 and 4 work. This fact limits the usefulness of year 5 course marks in providing concurrent placement validity evidence for CMPT. As for A2-Calculus, Table 6 suggests that scores on this section are related as strongly to marks in the other year 5 mathematics courses as they are to those in year 5 calculus. It should be noted, however, that the students taking CMPT in the COU administration had for the most part taken at least two year 5 mathematics courses and were therefore likely to be similarly prepared in these two areas. If, furthermore, we assume that mathematical ability is not narrowly confined by mathematical subject matter, then year 5 mathematics course marks should be highly intercorrelated. Thus an achievement test designed to assess the level of achievement in one course, calculus for example, would tend to correlate equally highly with marks in other

---

[14] Vincent D'Oyley et al, The Canadian Mathematics Placement Test (CMPT) Technical Report (August, 1976) available from COU on request.

mathematics courses. In this regard, it is worth noting that intercorrelations among year 5 mathematics marks in the 1975 COU sample ranged from 0.75 to 0.79. With due allowance being made for the less than perfect reliabilities of such single course measures, it is clear that the intercorrelations among "true" (without errors of measurement) course marks must be very high indeed.

Final first year university mathematics course marks are of little use in estimating the validity of CMPT for assessing the level of preparation, in various areas of mathematics, of students entering one year earlier. Similarly the estimation of CMPT validity for exemption purposes requires that testing take place concurrently with final examinations in the course concerned. Thus, the assessment of CMPT subtest validities for exempting students from normal course work requirements must rest on analysis of test content and its perceived adequacy in covering the objectives of the course to be exempted.

A wide variety of placement decisions fall into the c) category and the appropriate test validation procedures are intimately related to the specific decisions and to the instructional strategies in the institutions where the tests are to be used. It is impossible to do justice to the range and complexity of such decisions in a report of this kind.[15]

---

[15] For a full treatment the reader is referred to: Warren W. Willingham, College Placement and Exemption, College Examination Board (New York, 1974)

In practice, guidance and sectioning decisions are often based upon predicting the likelihood of a student's success or failure in courses or programmes which are more or less demanding of particular skills. If this is made the basis for a placement decision, then as Table 13.3 shows, some CMPT subtests and composites provide a good deal of information about subsequent grades in science, mathematics and engineering programmes. Using arguments analogous to those described under "Selection", the validity of CMPT for classifying students as "promising", "weak" (or whatever categories are chosen) can be estimated from the appropriate predictive validities. It should be noted, however, that where high school average is available, the CMPT adds little to prediction. Its principal use, then, would be in assessing likely success of students in mathematically demanding programmes where suitable Ontario high school data, based principally on mathematics and science achievement, is unavailable or suspect.

4.4 Variations in High School Marking Standards

4.4.1 The Problem

As was pointed out in earlier discussion, the variation among high schools in the marks assigned to students for "equivalent" achievement in year 5 courses compounds the already difficult classification decisions facing university admissions and guidance officers. Their principal concern in this regard is with equitable admissions decisions for students competing for places in limited enrolment programmes.

That high schools should have different marking standards is hardly surprising in view of the absence - since the departmental examinations were discontinued - of a common standard with which to relate. The presence of these differences does not in itself imply any deliberate manipulation of marks by some school authorities in order to enhance their graduates' chances in competitive admissions. Indeed the Ontario universities themselves do not share, nor should they be expected to share, common marking standards, even between similar programmes.

4.4.2 Methodology

In order to compare the marking standards of various high schools it is necessary to relate the marks they assign to some independent measure which is also a valid substitute for school marks.

One possible set of candidates for this role would be standardized tests. These will certainly provide scores which are "fair" with respect to the test-takers school of origin. On the other hand, they seldom show high concurrent validities with high school marks or high predictive validities with university marks to equal their high school equivalents (perhaps because they are a different "form" of evaluation compared to these two criteria, and because schools differ in the degree to which test content has been covered in their curricula). Thus even with their biases high school marks are statistically more useful to universities in assessing subsequent academic success than are standardized test scores.

Another approach is that of using a university measure as the independent standard for comparing high school marks. The measure often adopted is the mid-year or end of year programme average. An attempt is made to adjust student marks in accordance with the past performance of students coming from the same school. This is by no means an absolute standard for comparing high school marking standards since different programmes would reward different abilities and use their own scales to do so. The variability in measured university academic performance for students from the same high school and with similar high school marks makes the estimation of school differences a statistical exercise of some sophistication.

Adjustment procedures which take no account of the sample size upon which parameter estimates are based will likely lead to statistically unwarranted corrections to the marks of students from some schools, particularly from schools which supply small numbers of students to the programme from year to year.

The design of the University Records Survey in the Interface Study called for a stratified random sampling of schools throughout Ontario and of students within those schools. Thus findings concerning the estimated distribution of differences in school marking standards could be generalized to the population of Ontario high schools. No such generalization was possible with the COU samples, and thus there was nothing to be gained by attempting to include in the analyses every school and every participating university. Most of these schools supplied only a handful of students to the group of participating universities in each administration.

One of the primary objectives for a school variation analysis, that of investigating the possibility of reducing school biases by combining year 5 and achievement test scores, is not entirely ruled out by restricting the number of schools although it makes the analysis a case study rather than a general analysis. It was decided therefore to restrict analyses to a single university and to schools supplying significant numbers of students to each major first year programme.

McMaster University was the obvious choice since not only were there several feeder high schools supplying large numbers of first year students, but also data were available for this institution for two consecutive years so that analyses of the stability of prediction were possible.

The method of analysis adopted here to examine the effects of combining achievement tests with high school measures in order to reduce the effects of differences in high school marking standards, is essentially the same as that used and described in some detail in the University Records Survey.[16]

Even though several schools supplied substantial numbers of students to first year in each particpating university, the number of students from each school who registered in any particular university programme and who wrote a particular achievement test was (except for three schools) too small to provide stable estimates of parameters of interest in the analysis of school differences in marking standards. The difficulty was overcome by employing the method of multiple linear regression which allowed the university marks of all students from many high schools and registered in several different programmes to be related to their high school marks and test scores, in a unified analysis. Some measure of a student's university academic performance, such as first year programme average or first year mathematics average, is considered to be the dependent variable which is to be "predicted" from such information

---

[16] Ibid; pp. 156 - 164

as the student's high school marks, test scores, university programme and secondary school attended. The entry of the latter two types of predictor information into the analysis has the effect of "adjusting" for any differences in the marking standards employed by the various evaluation units. The adjustment estimated for each secondary school or university programme is such as to bring predicted first year marks closely into line with those observed in the student sample as a whole.

The school regression coefficients which result from the multiple regression analysis summarized in Tables 17.1 to 18.2 are to be interpreted as the marks which must be added to a student's predicted first year programme average, in order to adjust for variations in the performance of a student from a particular secondary school. In each table, separate analyses A and B are presented to show the magnitudes of school regression coefficients when predictors are based on high school marks only (A), and when a composite of standardized test score and high school mark (B) is used. Since the effect of year 5 subject or overall average is taken into account in analyses A. and B. of these tables, the most likely explanation for differences in school regression coefficients is that they represent differences in the marking standards employed by the high schools (as reflected in university academic performance one year later).

Table 17.1

Summary of Regression Analyses of First Year Average (1975-76) Prediction for McMaster Registrants from Ten High Schools who wrote CMPT-1

		Predictors in Order of Entry into Prediction Equation	Multiple R Sample	Adjusted	F-Ratio	d.f.
A.	1.	Year 5 Average	0.51	0.51	41.07	1:115
	2.	University Programme	0.55	0.53	2.99	2:113
	3.	Secondary School	0.75	0.71	6.80	9:104
B.	1.	Year 5 Average & CMPT-1	0.55	0.53	24.16	2:114
	2.	University Programme	0.60	0.58	5.39	2:112
	3.	Secondary School	0.76	0.73	6.05	9:103

Estimated School Regression Coefficients

School	A. School Regression Coefficient	S.E. of Coefficient	B. School Regression Coefficient	S.E. of Coefficient
1.	-6.3	2.4	-4.9	2.4
2.	5.9	2.3	4.3	2.3
3.	1.5	2.0	1.2	1.9
4.	4.1	3.2	3.7	3.1
5.	1.2	4.5	2.7	4.5
6.	-27.0	4.1	-26.3	4.1
7.	3.8	3.9	4.7	3.8
8.	14.7	5.2	14.5	5.1
9.	3.4	3.8	3.8	3.7
10.	-1.3	-	-3.7	-

Table 17.2

Summary of Regression Analyses of First Year Mathematics Average (1975-76) Prediction for the McMaster Registrants from Ten High Schools who wrote CMPT-1 (Algebra)

Predictors in Order of Entry into Prediction Equation	Multiple R		F-Ratio	d.f.
	Sample	Adjusted		
A. 1, Year 5 Mathematics Average	0.45	0.45	29.80	1:115
2. Secondary School Attended	0.65	0.60	4.36	9:106
B. 1. Year 5 Mathematics Average & CMPT-1 (Algebra)	0.52	0.50	20.79	2:114
2. Secondary School Attended	0.67	0.62	3.82	9:105

Estimated School Regression Coefficients

School	A. School Regression Coefficient	S.E. of Coefficient	B. School Regression Coefficient	S.E. of Coefficient
1.	-4.6	3.6	-1.5	3.8
2.	8.1	3.4	5.1	3.5
3.	0.1	3.0	-1.2	3.0
4.	9.7	4.7	8.8	4.6
5.	7.7	6.7	9.9	6.7
6.	-28.7	6.1	-27.3	6.0
7.	-3.4	5.6	-2.7	5.5
8.	19.7	7.7	20.2	7.5
9.	-1.9	5.6	-0.7	5.5
10.	-6.7	-	-10.6	-

Table 18.1

Summary of Regression Analyses of First Year Average (1976-77) Prediction for McMaster Registrants from Ten High Schools who wrote RCLAT

	Predictors in Order of Entry into Prediction Equation	Multiple R Sample	Adjusted	F-Ratio	d.f.
A.	1. Year 5 Average	0.51	0.51	238.51	1:666
	2. University Programme	0.60	0.59	23.25	4:662
	3. Secondary School	0.65	0.64	8.86	9:653
B.	1. Year 5 Average & RCLAT Composite	0.53	0.52	63.88	4:663
	2. University Programme	0.60	0.60	21.78	4:659
	3. Secondary School	0.65	0.64	8.05	9:650

Estimated School Regression Coefficients

School	A. School Regression Coefficient	S.E. of Coefficient	B. School Regression Coefficient	S.E. of Coefficient
1.	-2.3	1.1	-2.1	1.1
2.	5.0	0.8	4.9	0.8
3.	1.6	0.8	1.6	0.8
4.	1.2	1.4	1.4	1.4
5.	2.1	1.5	2.1	1.5
6.	-8.5	1.5	-8.4	1.5
7.	2.2	1.3	2.1	1.4
8.	-2.2	1.5	-2.1	1.5
9.	-1.3	1.8	-1.4	1.8
10.	2.2	-	1.9	-

Table 18.2

Summary of Regression Analyses of First Year English Average (1976-77) Prediction for McMaster Registrants from Ten High Schools who wrote RCLAT

	Predictors in Order of Entry into Prediction Equation	Multiple R			
		Sample	Adjusted	F-Ratio	d.f.
A.	1. Year 5 English	0.34	0.33	22.76	1:177
	2. Secondary School	0.53	0.49	4.32	9:166
B.	1. Year 5 English & RCLAT Composite	0.44	0.42	10.41	4:174
	2. Secondary School	0.58	0.53	3.80	9:165

Estimated School Regression Coefficients

School	A. School Regression Coefficient	S.E. of Coefficient	B. School Regression Coefficient	S.E. of Coefficient
1.	3.6	2.6	5.7	2.7
2.	5.5	1.7	3.9	1.7
3.	4.3	1.7	3.5	1.7
4.	-4.1	2.9	-0.8	3.0
5.	3.6	2.9	2.4	2.9
6.	-15.1	3.6	-13.8	3.6
7.	4.1	2.7	2.1	2.8
8.	-0.7	3.2	-1.1	3.2
9.	-9.4	4.7	-8.3	4.7
10.	8.2	-	2.4	-

In addition to school regression coefficients and their standard errors, Tables 17.1 to 18.2 also contain information regarding the incremental effectiveness of the high school and achievement test predictors in accounting for first year university performance. The multiple correlation between predictors entered up to that point and first year average, the F-ratio to test the significance of the contribution made by each predictor, and the degrees of freedom associated with this F-ratio are shown on each table. In addition to the multiple correlation found in the sample, a more conservative estimate is provided which adjusts for the number of independent variables and the number of students in the sample, and more accurately reflects the strength of the relationship between the criterion and predictor variables that would be expected to hold in the hypothetical population from which the students are considered a sample.

4.4.3 Results

The ten schools were selected solely for the size of their contribution to the first year intake at McMaster and are consequently not in any way representative of any wider population of schools. Thus the multiple correlations presented in Tables 17.1 to 18.2 are not to be taken as applying to the COU sample as a whole. Unfortunately, the numbers of students coming from schools 4 through 10 were quite small and some unrepresentative results by a few students could have been responsible for the larger deviations from the average for all schools. The data are nevertheless presented, in that they are statistically correct and are included for the sake of completeness.

The principal interest in these tables lies in:

a) the significance of the contribution of "secondary school attended", to the prediction of university performance, before and after a standardized measure has been entered into the prediction equation; in all cases the contribution as determined from the F-ratio provided is significant (at $p>.01$);

b) the school regression coefficients, their standard errors and the change in their values when a standardized predictor is combined in a composite with the unstandardized high school predictor. The standard error of the school regression coefficient provides a basis for judging whether the coefficient differs from zero by chance, i.e. as a result of sampling fluctuations. If the coefficient is more than twice its standard error, then the probability is less than one in twenty that the school sample observed comes from a school having a school regression coefficient of zero, i.e. a school for which no adjustment to the predicted university performance of its graduates is justified.

For most of the schools, on each table, the school coefficients cannot be distinguished from zero according to the criterion given above. Some schools, however, are consistently higher or lower than zero and significantly so. Students from School 2, for example, consistently obtain better university results in English, mathematics and overall programme average than would be predicted from their high school marks alone (and

programme of study for programme average predictions). School 6, on the other hand, has a school regression coefficient which is consistently much lower than the average for all schools, by an amount which varies from 8.5 in Table 18.1 to 28.7 in Table 17.2. The regression coefficients for other schools are less consistent from table to table and, for the most part, are not significantly different from zero.

The effect of combining a standardized test score with the high school measure can be seen in Column B of each table. There are, for several schools, reductions in the absolute size of school regression coefficients, but they remain substantial. Since the bias of a composite predictor depends on the relative weights given to the biased and the standardized components, an extra weighting on the latter component will reduce the bias but also reduce the predictive validity of the composite from the maximum value obtained by using the weights estimated in the regression analysis.

It would appear then that, for the ten schools studied, the addition of a <u>single</u> standardized test score to high school predictors would not correct, to any great extent, for the differences in the marking scales employed by high schools.

5.0 SUMMARY OF FINDINGS

A major objective of the two COU projects reported here was that of getting some hard empirical data on the utility of various measures of pre-university academic achievements and aptitudes which are routinely used, or might be so used, in policy, admissions and in post-admission decisions. The primary focus in these projects was upon what additional information could be provided by the use of standardized achievement tests. Studies such as this cannot provide all the answers to the possible questions that are apt to be asked. Limitations of resources, suitable data and sampling design restrict the ability to draw firm conclusions in many areas. The following is a list of potential aims of research studies on achievement testing;

- estimating aggregate trends in achievement levels in areas such as English language and mathematics;
- providing benchmarks in aggregate achievement levels in these areas;
- obtaining experience in large scale test administration;
- obtaining useful data for the development and refinement of promising tests;
- assessing test utilities for operational admissions and post-admissions decision-making by or for individual students;

- investigating the possible uses of achievement tests in standardizing or replacing high school achievement measures.

Before summarizing findings in the above categories some general comments seem appropriate.

First, we have found a dearth of "genuine" criterion measures, certainly none with indisputable credentials, for estimating the empirical validities of test measures for some uses. On a practical level comparable findings between institutions can rarely be made with confidence because of variations in definitions and record keeping practices.

The findings of the COU achievement testing projects are summarized below in the order outlined previously.

a) Aggregate Trends

The lack of comparable reference samples on the tests administered in the COU projects precludes any findings of note in this area.

b) Benchmarks

Benchmarks for all the tests used are available for the participating institutions. This remark must be qualified by observing that in the 1975 administrations, for CTEL

and CMPT, the tests were not mandatory and some non-random self-selection is evident in the volunteer sample. Thus the benchmarks obtained for the institutions participating in the experimental testing for CTEL and CMPT are not truly representative of the 1975 first year freshmen cohort. The usefulness of these levels for the participating universities will depend on whether the same or equivalent tests are used again in future. Recommendations for the modification of the tests as a result of the experimental administration results, if implemented, would require that the norms reported here be adjusted by equating scores through an experimental administration of both old and new forms, before they could be used for aggregate comparisons.

c) Test Administration Experience

Valuable experience of large scale testing was gained in the course of the two testing programmes by the participating universities and COU research staff.

d) Test Development

For each of the tests used in the COU administrations, results have brought to light areas for improvement in test quality. The recommendations can be found in the seperate technical manuals.

e) Test Quality

The tests, as administered, were generally found to possess satisfactory internal qualities. Low reliabilities, however, on the RCLAT subtests and the Writing Test limit their independent usefulness for some purposes.

Concurrent and predictive validities of the English test scores from CTEL, RCLAT and the Writing Test with year 5 English and university English averages as criteria, are in the range commonly found in validation studies of achievement tests scores for university admissions. Estimated concurrent validity with year 5 English was 0.49 for CTEL total score and 0.46 for RCLAT total. Predictive validities with university English averages for these same measures were estimated from 0.33 to 0.41 for CTEL total score and 0.34 and 0.38 for RCLAT total score.

Although Writing Test scores are almost certainly less reliable than those of either of the English language objective tests, correlational results suggest that the Writing Test has higher validities against year 5 and university English criterion measures and discriminates better between English and mathematics criteria.

The estimated concurrent validities of CMPT subtest scores with year 5 mathematics average varied from 0.46 to 0.58. Predictive validities, depending on subtest and university, lay between 0.22 and 0.60 with a median of 0.41 using first year final mathematics average as criterion measure. If the criterion is further refined by restricting marks used to those of students undertaking a substantial course load in university mathematics then the predictive validities of CMPT subtest scores fall in the range 0.37 to 0.68 with a median of 0.52.

f) Achievement Test Validities for Student Classification Decisions

A number of analyses were carried out on the COU data which had implications for some of the possible uses of achievement test scores by universities. Taking each of the principal uses in turn, the findings are as follows:

i) Selection for limited enrollment programmes:

In general, where appropriate year 5 information was available, no test score or combination of test scores (at least for such as were available for analysis) effected more than a small gain in selection efficiency at selection ratios likely to be encountered at the present time in Ontario. Failing the availability of year 5 data however, two CMPT subtests consistently showed moderately high levels of predictive validity for science, mathematics and engineering programmes

albeit with some reduction in selection efficiency. CTEL and RCLAT scores were much less useful for programme selection than the other available measures. The relative benefits of correcting for variations in high school marking scales could not be adequately estimated from the sampling design employed in the COU studies.

ii) Placement

Estimating criterion related validity of the achievement tests for placement purposes was precluded by the lack of suitable criterion measures and experimental design. Under the "placement" rubric we include the assignment of students to remedial, regular or advanced English courses. None of the English achievement tests administered by COU in 1975 or 1976 have been properly validated for this purpose e.g. does remedial English instruction make any substantial difference to the ultimate development of the criterion English language skills and do the tests efficiently select those students who will benefit the most from remedial instruction?

iii) Standardization of High School Marks

Although the addition of standardized test score information to year 5 marks information effects some reduction in the bias attributable to differences in high school marking scales, it would seem on the basis of results from 10 high schools that a substantial bias still remains in the composite selection measure.

COU 1975: Achievement Test Volunteers by University and Final Programme of Registration

Programme	Brock			McMaster			Queen's			Waterloo		
	N_P	N_{CTEL}	N_{CMPT}	N_P	N_{CTEL}	N_{CMPT}	N_P	N_{CTEL}	N_{CMPT}	N_P	N_{CTEL}	N_{CMPT}
Arts	417	332					850	148		814	303	19
Humanities				343	193	3						
Social Sciences				797	328	8						
Natural Sciences	146	120		678	399	390	526			561	257	252
Mathematics										838	511	507
Engineering				400	270	271	425		143	716	361	363
Environmental Studies										394	188	39
Commerce & Business Administration	97	78		441	224	219	175		111			
Physical Education & Health Sciences	126	109		370	143	6	90			315	114	70
Nursing				72	46	1	60					
Occupational & Physical Therapy							47					
Other				81	31	1	79			18	1	
Missing Programme Information				58	27	17	224		1	155	69	50
TOTAL	786	639		3240	1661	916	2476	148	255	3811	1804	1300

COU 1975: Educational Source of Registrants

Source	Brock		McMaster			Queen's			Waterloo		
	N	N_{CTEL}	N	N_{CTEL}	N_{CMPT}	N	N_{CTEL}	N_{CMPT}	N	N_{CTEL}	N_{CMPT}
Ontario Year 5 (Current)	384	339	2352	1501	848	1678	122	205	3013	1573	1160
Ontario Year 5 (Prior Year) and Preparatory*	76	65	203	77	30	104	7	9	227	82	38
Year 5 Equivalent Other Canadian Province	6	3	16	4	3	125	8	18	52	28	23
Year 5 Equivalent Other Country	39	22	66	21	11	79	4	6	72	34	20
Transfers from Post-Secondary Source	112	72	148	28	13	123	6	14	197	43	35
Miscellaneous**	167	138	103	30	11	26	1	3	53	44	24
Missing Source Information			352			276			197		
Total	786	639	3240	1661	916	2911	148	255	3811	1804	1300

158

* Preparatory Studies: Ontario Correspondence School, Night School, preliminary year at Ontario university

** Miscellaneous: mainly Ontario Year 4 and adult students

159

COU 1976: Achievement Test Writers by University and Programme of Registration

Programme	Carleton		McMaster		
	N_p	N_{rclat}	N_p	N_{rclat}	N_{essay}
Arts:					
Humanities	1361	1361	363	311	100
Social Science			848	628	186
Natural Science	317	317	664	576	171
Engineering	241	241	326	294	86
Industrial Design	41	41			
Architecture	61	61			
Business/Commerce	211	211	441	375	89
Physical Education			242	211	57
Nursing			70	63	23
Journalism	168	168			
Other	8	8	62*	41	8
Missing Programme Information	134	133	30	30	10
Total	2542	2541	3046	2529	730

* Includes Music and Experimental I

COU 1976: Achievement Test Writers by Educational Source

Educational Source	Carleton N_s	Carleton N_{rclat}	McMaster N_s	McMaster N_{rclat}	McMaster N_{essay}
Ontario Year 5 (Current)	1560	1560	2354	2134	637
Ontario Year 5 (Prior year)	164	164	159	116	28
Preparatory*	17	17	11	8	2
Other Canadian Province (Year 5 Equivalent)	64	64	10	10	3
Other Country (Year 5 Equivalent)	12	12	40	29	4
Post-Secondary (CAATS or Equivalent)	117	117	85	52	10
Post-Secondary (Universities)	67	67	174	21	3
Ontario Year 4	29	29	30	19	8
Missing Source Information	512	512	183	140	35
Total	2542	2542	3046	2529	730

Notes on Selection Efficiency

For a given programme and selection measure we define:

- the selection ratio as the ratio of the number of students selected from the applicant pool to the total number of applicants;

- the success ratio as the proportion of students selected who were subsequently successful in the programme concerned;

- the selection efficiency as the proportion of students selected at a selection ratio of p who were also among the highest ranked p on the criterion measure of programme success.

If it be assumed that in selecting for admission to limited enrollment programmes, the best possible selection procedure would be to select the highest ranked students on the university evaluated criterion measure (e.g. first year programme average) in numbers sufficient to fill the available places, then the selection efficiency as defined above, is the appropriate index for comparing the relative merits of different selection measures.

The selection efficiency varies both with the selection ratio and with the predictive validity of the measure used in selection. This variation is

displayed in Table A5 for the case where the joint distribution of selection measure scores and criterion measure scores is approximately bivariate normal.

The predictive validity estimated from a sample should, if possible, be corrected for criterion unreliability, since the "best" possible selection measure would be the "true" criterion score rather than the observed criterion score with its error component.

An example may serve to elucidate the use of Table A5. Consider a programme having 1000 places to be filled and 2000 applicants. From the definition above the selection ratio for this programme would then be 0.50. If analysis in previous years had established that the correlation of student year 5 averages with first year university programme averages was 0.70 (when the observed correlation is corrected for the effects of direct selection on the year 5 average measure, and for the unreliability of first year average) consulting Table A5 at r = 0.70 and selection ratio = 0.50 gives the entry 0.75. Thus we would expect to select 750 of the "best" 1000 applicants for admission to the programme when the year 5 measure is used in selection. The other 250 selected by this measure would not have ranked in the first 1000 applicants had the first year average been used in selection.

Table A5

Selection Efficiencies for Various Values of the Selection Ratio and the Predictive Validity (r), of the Selection Measure

	Selection Ratio										
r	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95
.00	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95
.05	.06	.12	.22	.32	.42	.52	.61	.71	.81	.90	.95
.10	.07	.13	.24	.34	.44	.53	.63	.72	.81	.90	.95
.15	.09	.15	.26	.36	.46	.55	.64	.73	.82	.91	.95
.20	.11	.17	.28	.38	.48	.56	.65	.74	.82	.91	.95
.25	.12	.19	.31	.41	.49	.58	.66	.75	.83	.91	.95
.30	.14	.22	.33	.43	.51	.60	.68	.75	.83	.91	.95
.35	.17	.24	.36	.45	.53	.61	.69	.76	.84	.92	.96
.40	.19	.27	.38	.47	.56	.63	.70	.77	.85	.92	.96
.45	.22	.29	.41	.50	.58	.65	.72	.78	.85	.92	.96
.50	.24	.32	.44	.52	.60	.67	.73	.80	.86	.92	.96
.55	.28	.36	.47	.55	.62	.69	.75	.81	.87	.93	.96
.60	.31	.39	.50	.58	.64	.70	.76	.82	.87	.93	.96
.65	.35	.43	.53	.60	.67	.73	.78	.83	.88	.94	.97
.70	.39	.47	.56	.63	.69	.75	.80	.84	.89	.94	.97
.75	.44	.51	.60	.67	.72	.77	.81	.86	.90	.95	.97
.80	.50	.56	.64	.70	.75	.80	.83	.87	.91	.95	.97
.85	.56	.62	.69	.74	.79	.82	.86	.89	.92	.96	.98
.90	.64	.69	.75	.79	.82	.86	.88	.91	.94	.97	.98
.95	.73	.78	.82	.85	.87	.90	.92	.94	.96	.98	.99
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Source: Abstracted from appropriate entries in: H.C. Taylor & J.T. Russel Journal of Applied Psychology Vol. 23 pp. 565 - 578